Bachelor Graduation Project Tracheo-esophageal speech enhancement:

Real-time pitch shift and output

July 23, 2020

TU Delft

Group I-2:
Tim Alers, 4585771
Bosse Fennema, 4590635
Joris van Breukelen, 4704967

Abstract

Laryngectomised people (LP) have many problems with speaking in terms of intelligibility, volume and effort it takes. This thesis describes the design and simulation of a voice enhancement device which will improve the speech of LP with respect to intelligibility and volume. It only describes the chosen hardware peripherals in order to record, process, amplify, and output the modified speech. The recorded audio could be processed using a digital signal processor (DSP) that could be programmed such that it will filter the undesired noise and shift the pitch of the speaker's voice in real-time. This thesis focuses on the real-time pitch shifting of the voice, the hardware choices and implementation of the sound output of the system, and the implementation of DSP of the Tracheo-Esophageal Speech Signal Amplifier (TESSA). The real-time pitch shifting makes use of the Ocean algorithm, of which the principle is described, implemented and tested. Intelligibility tests are performed in order to find the optimal shifting parameters.

Keywords— Laryngectomised people, Speech processing, Digital signal processor, Real-time pitch shifting, Ocean Algorithm, Intelligibility

Preface

This thesis describes the early development stages of a new product, the TESSA (Tracheo-Esophageal Speech Signal Amplifier). This device should help those who lost their voice due to a laryngectomy, regain the ability to speak and be heard. This document focuses on the pitch and loudness of the voice, and how to improve them in real time. It is written in partial fulfilment of the Electrical Engineering Bachelor's programme of the Delft University of Technology. The thesis is written as part of a larger project. The other subsystem is designed and documented by Johan Meyer and Folkert de Ronde. During the project we have helped each other and exchanged ideas on a daily basis.

The project was requested by Guus van Wechem, an alumnus who after his own laryngectomy decided he was not satisfied with what the current market had to offer. He has set out to create a new product, the Exo-breather, to better suit the needs of all laryngectomy patients. This project, which started in the month of April 2020, is one of the first steps to realising this goal and will lay a foundation for future researchers to base their work on. The circumstances of writing this thesis are cumbersome, as the world is in lock-down due to the corona virus outbreak. We are not allowed to meet as a group and physical tests or a real-world demonstrator cannot be part of our thesis. Instead of building a prototype for the product, we use arguments based in theory and simulations to make recommendations for the next stages of development.

We would like to thank those who have helped us create this thesis, starting with Guus and his companion Peter. They have not only given us the opportunity to be a part of this project, but were also always very open and enthusiastic when we were exchanging ideas. Furthermore, we would like to thank our supervisors Ron van Puffelen and Jeroen Bastemeijer, without whose guidance and counsel this thesis would not have been possible. Lastly, we appreciate the help of the experts in the field of speech therapy Klaske van Sluis and Rob van Son, and the professors Richard Hendriks, Gerard Janssen and Alle-Jan van der Veen for their input during our design process.

Contents

1 | Introduction

Our voice is an essential part of our daily life. It is used to communicate with others and a life without proper communication is hard to imagine. Laryngectomised people (LP) still do not have a ideal solution to their problem of defective speech. Laryngectomised people have undergone a laryngectomy, which is the removal of the larynx and voice box due to medical issues considering these organs, shown in Figure 1.1. Most of these patients have a trach to enclose the hole in their throat near the original place of the larynx to regulate airflow in and out of the lungs. LP have big problems with speaking since it is significantly less intelligible, does not sound pleasant to listen to, and it requires much more energy which is very exhausting for the patient. The goal of this project is to design a device which will increase the intelligibility, voice quality and sound level of the LP. This thesis focuses on the pitch shifting part of the TESSA and its sound output. However, the entire project is done during the corona outbreak in 2020 in the Netherlands. This obstructed the physical implementation of the system and therefore this thesis will only be a theoretical description of the design of the device. Using simulations on hardware and by testing the speech processing using computer software the thesis will be as realistic as possible.

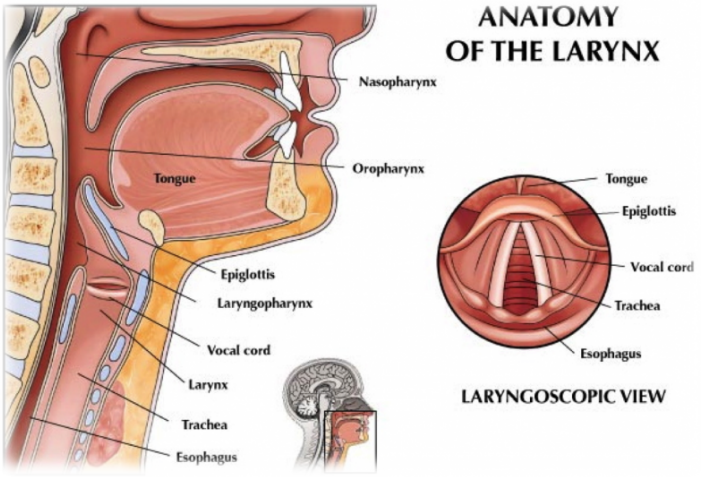


Figure 1.1: Schematic of the anatomy of the Larynx in a healthy person [23]

1.1. Speech characteristics

Human speech is a complex system with a lot of characteristics which differ from person to person. Three important characteristics of a voice are the fundamental frequency (F_0), the jitter and the shimmer. The fundamental frequency is a very important frequency of a voice and determines how high the pitch of the voice is. The jitter is the percentage difference in period between two consecutive waves of in the sound of the voice and the shimmer is the percentage difference in amplitude between two oscillations [7]. F_0 can be determined and shifted, which shifts the harmonics as well, changing the pitch of the voice. The jitter and the shimmer are for now only used diagnostically and modifying them is not in the scope of this project.

Normal speech

The speech characteristics for normal speech are considered for comparison with the laryngectomised speech. A teacher will speak around 2 hours effectively a day according to other research [39]. This will be used as reference for maximum amount of speaking time which is needed for the device to be able operate in terms of energy consumption. Furthermore the functionality of the normal speech is that the fundamental frequency is filtered and amplified by the windpipe and mouth. These modifications make it so that the fundamental frequency in combination with the vibrations of the throat results in an audible sound.
The frequency range of the voice is between 500 Hz and 4 kHz. However the intelligibility is mostly reliant on the 2 – 4 kHz range [30], which is important for one the desired improvement objectives of the device.

Laryngectomised speech compared to normal speech

Two key aspects in which laryngectomised speech differs from normal speech are the pitch and loudness of the voice. An LP typically speaks with a much lower fundamental frequency, because their method of voice production is not able to replicate the high frequencies of the vocal chords. The F0 of a male LP might range from 50-110 Hz, whereas a male or female with vocal chords have a range of 85-180 Hz or 165-255 Hz respectively. The difference could make an LP voice sound unnatural, especially for a woman [32].
The loudness of one's voice is very important in everyday life, although people might not realise it. For LP it is not a given to be able to have a conversation in a noisy environment due to the low volume they produce. For soft and loud speech, LP have been found to produce sound levels close to standard [5]. In this research loud speech was defined at 60 dB. But when raising one's voice, a healthy person can reach levels up to 85 dB to make oneself heard in a noisy environment [9].

1.2. Current reinstatement methods

There are currently a few ways in which it is possible for LP to have an improved speech. The Electrolarynx is an electrical device that induces vibrations in the the throat to enable speech. It is a general purpose device that can be used by any patient, but it requires a handheld device to be put on the throat or in the mouth while speaking. This makes it very visible and uncomfortable for the user. Furthermore, the sound at a constant fundamental frequency sounds very robotic and retains little of the user's original voice. The lack of variation in frequency is also found to decrease the intelligibility of the produced speech [27].
A patient can also swallow little portions of air and let the air out in a controlled manner to generate speech, this is called esophageal speech [24]. Another method for speech reinstatement is by having an operation by which the oesophagus is connected to the windpipe by a valve. The valve can be used to let air from the windpipe flow to the oesophagus and then into the mouth. This method is called tracheo-esophageal speech [41]. The speaker must train to create vibrations using the tissue in the throat. With the vibrating air, speech is formed. Even with years of practice, these speech methods do not reach the same level of acceptability or intelligibility as regular speech although tracheo-esophageal speech does yield better results [40].

1.3. The system

The TESSA is made in regard of the tracheo-esophageal speech. Meaning that it will only work for people who have had an tracheo-esophageal operation. The system will record the tracheo-esophageal speech, process the speech and output the amplified and modified audio in real-time. Therefore the system will consists of a microphone for recording, a digital signal processor (DSP), a speaker together with an amplifier, and a battery to power the entire system. It is designed to be part of the Exobreather, which is a device worn on the body that will also regulate heat and moisture in the user's oral cavities.

1.4. Structure of the thesis

This thesis will start by giving a clear design overview of the entire system. Since this thesis describes a sub-part of the entire system, it will elaborate more on the overview of the design of the sub-part. Following chapters will provide the theoretical description of the parts of the system this thesis is focused on. These are the shifting algorithm needed in the speech processing and the hardware peripherals for the sound output of the system and the digital signal processor. Chapter 5 will describe additional features and the expected performance of the device. This is followed by a theoretical implementation and testing phase, which is done using simulations and using MATLAB for the speech processing. Chapter 8 and 9 will give a discussion and conclusion respectively. Afterwards, there will be recommendations for follow-up research.

2 | Design considerations

To align the expectations of the designers and the client, a Statement of Requirements (SOR) was agreed upon and can be found in Appendix G. The subsystem of the voice enhancement product described in this thesis handles the shifting of the signal and the output.

2.1. General requirements

The essential selection of requirements for this subsystem is stated below:

- The System comprises:
  - A signal processing part
  - An amplifier
  - An actuator
- It is wearable on the body
- The input bandwidth is at least 50 Hz to 4 kHz;
- The maximum output sound intensity is at least 85 dB at 0.3 m;
- The output bandwidth is at least 50 Hz to 4 kHz;
- Its operating temperature stays below $36°C$;
- Its power consumption does not exceed 2.25 W.

The requirements stated here are partly derived from the SOR, and partly from the conditions set by the other subsystem. The analogue speech data is sampled at 16 kHz with 16 bits per sample in the first subsystem. It is then filtered and the output of the final filter is the input of the subsystem described in this thesis. The output of the filtering subsystem is a filtered spectrum with the bandwidth of 8 kHz. Only the necessary information of the speech signal are received at the input of the shifting subsystem. What remains is to shift and amplify the signal to make it sound more natural and intelligible before sending it to a playback device.
The intended sound intensity of 85 dB at 0.3 m at the output will be enough for the user to make him- or herself heard in noisy environments. To keep the amplified voice intelligible the bandwidth containing the frequencies of importance in human speech have to be amplified. This bandwidth is found to be 500 Hz to 4 kHz [30]. In the interest of being able to include the fundamental frequency, the input and output bandwidth are chosen between 50 Hz and 4 kHz. As the product should be wearable on the body it should not heat up to become uncomfortable in use. Therefore the maximum operating temperature is chosen at $36°C$ which mostly affects the necessary efficiency of all components to keep heat dissipation limited [6].
The other subsystem contains the power source that also supplies this subsystem. The energy budget is chosen such that the shifting and output can have a power consumption of maximally 2.25 W. This includes the amplifier and the actuator but also the data processing device. For the amplifier, actuator and data processor off the shelf components will be selected after an extensive selection procedure. The created system must comply to the above requirements. The volume regulation of the system should be managed hands free.

2.2. Signal processing requirements

Additional requirements are set specifically for the necessary signal processing:

- The speech signal should be shifted up and sound more natural as a result;
- The voice should be intelligible.

2.3. Preferences

Another set of additional preferences were set up to guide the design process:

- The size will have to be as limited as possible;
- The total costs of the parts should be around 100 euros;
- It is integratable with the 'Exobreather';
- The latency of sound produced should stay below 15 ms;
- The reproduced voice should be as natural as possible;
- The volume should be regulated by air pressure.

The above preferences are guidelines to keep in mind and are mostly topics of future research. These are for now unessential goals. For example the component price should stay near 100 euros to make the product affordable. And as the product records, processes and amplifies the same voice the produced sound of the product could be perceived as an echo due to latency. To prevent this echo from being audible and forming a real hinder the latency should stay below 15 ms. Preferably even below 10 ms [18]. Ideally the product can perform all above requirements and preferences without losing the natural touch of a voice. And ideally the volume is regulated using the air pressure generated by the LP as in a healthy voice the volume is also regulated by increasing the airflow passing the vocal chords by increasing the air pressure.

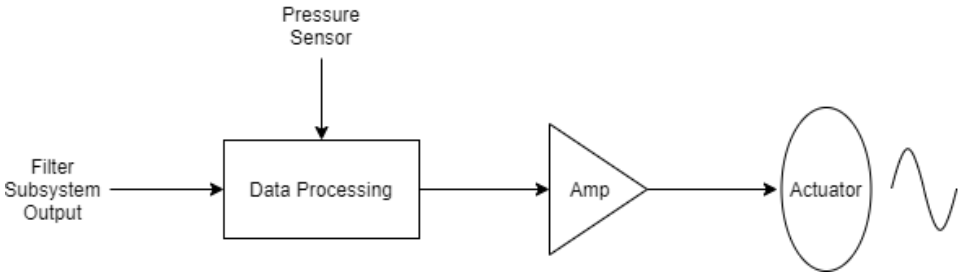


Figure 2.1: A low level block diagram showing a general overview of the subsystem.

The set of requirements and preferences are compiled into a preliminary high-level block diagram of the subsystem which is shown in Figure 2.1. In the following chapters, the method of shifting the signal and most optimal off-the-shelf hardware will be determined. Then the desired functionalities of each component is described in more detail. Finally, the integration of the subsystem is described and all components are put together.
Unfortunately, it will not be possible to build the system over the course of this thesis. Therefore there can be no testing in a real environment. Through simulations and tests in a digital environment, this thesis attempts to show that the design will meet the requirements as stated in this chapter.

3 | Shifting

3.1. Requirements

The purpose of shifting the LP voice in frequency domain is to increase the intelligibility of the LP voice and make it sound more natural. There are two ways of changing the pitch of a signal: frequency shifting or pitch shifting. Frequency shifting changes all the frequency components with a constant value, from example from 300 Hz to 330 Hz and 600 Hz to 630 Hz (an increment of 30 Hz). Pitch shifting changes the pitch with a constant factor, from example from 300 Hz to 330 Hz and 600 Hz to 660 Hz (an increment factor of 1.1). Changing the pitch of the human voice requires to preserve the ratios of the harmonic frequencies that determine the formants and therefore pitch shifting is the only suitable option [17].
Furthermore, the shifting has to be done real-time. Therefore the delay of the implementation should not exceed 15 ms. The pitch should be adjusted without changing the playback speed.

After all, the pitch shifting is a function that will be called by the filters. The input is a digital signal containing the LP voice in frequency domain and the required shifting factor. The background noise is already removed by filters. The bandwidth is limited to 8 kHz. The windowing has been applied in the filter implementation as well.

3.2. Implementation

There are several ways to implement a real-time pitch shifter. Existing pitch shifters are used in the music industry, for example in modulation effects and Autotune. Those pitch shifters can be found in effect pedals or in recording software. Both options are very expensive, require a lot of hardware and are very bulky for the exobreather application. However, it is possible to implement a pitch shifting algorithm on a digital signal processor (DSP). In consultation with the filter implementation, the latter is chosen. The selection of the DSP will be described in Section 4.1.

3.3. Algorithms

There are multiple implementations of shifting algorithms. The goal is to use an algorithm that retains the sound quality and intelligibility of speech while shifting. It should also be possible to run the algorithm in real-time on a DSP, and thus cannot be too computationally complex.
Each of these algorithms come with associated costs and benefits. Jan Onderka has done an analysis of the most prominent options [26]. This research also focused on choosing an algorithm to be implemented on a DSP. The Ocean Algorithm emerged as the most viable solution due to the balance between sound quality and low required computing power it provides. The Roller algorithm is also a notable mention, it does provide superior performance such as lower latency but at the cost of very large computing power demand. The Roller algorithm could become a more attractive solution when DSPs become more powerful. For this design, however, the Ocean algorithm is selected.

3.4. Ocean Algorithm

Principle

The Ocean algorithm operates using the Short Time Fourier Transform (STFT) of a signal [12]. The DFT is taken of a small number of samples at a time. This number will henceforth be referred to as the analysis frame size N_a. The resulting DFT has a magnitude at each frequency bin, corresponding to peaks in the frequency spectrum. The Ocean algorithm copies the content of each bin to a new bin that is a factor higher. Let a denote the original bin number and k the factor with which the frequency will be scaled. Equation 3.1 describes how the new bin is chosen.

$$b = \lfloor ka + 0.5 \rfloor \tag{3.1}$$

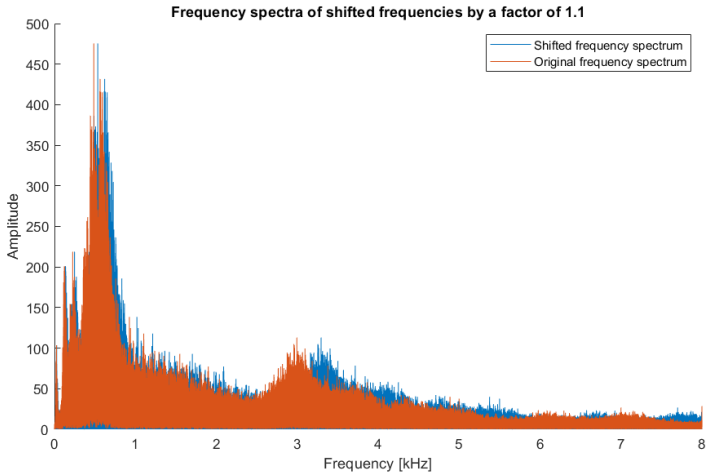


Figure 3.1: Frequency spectrum of the original and the shifted audio fragment

Implementation

The Ocean algorithm is implemented with MATLAB as function that takes in several parameters. The original signal is already cut into smaller frames of about 10 ms before shifting, because this is required for real-time pitch shifting. The original time domain signal is also converted to a frequency spectrum using the fast fourier transform (FFT) which is a faster and more efficient implementation of the DFT. Since the shifting module follows up the filter module, the original is already filtered as well. k is the factor of the scaled frequency spectrum, osamp is the overlapping variable, which is 1 in case of no overlap. A for loop is used to implement Formula 3.1. For every bin, the new bin is calculated. Because of the fact that the fast Fourier transform is symmetric, only half of the new bins have to be calculated, the other half is a mirrored version of the first half.

A phase correction is needed to make the signal propagate well between two frames. Formula 3.2 shows the phase correction which will be performed on all phasors on every frequency bin. Ω_b is the phasor on frequency bin b and will be multiplied by $e^{-i\frac{(b-a)p}{O}\frac{2\pi}{N}}$, which is dependent on the frame number and how much the frequency bin is shifted.

$$\Omega_b = \Omega_b e^{-i\frac{(b-a)p}{O}\frac{2\pi}{N}} \tag{3.2}$$

FFT size and sample frequency

In the publication of the Ocean algorithm, a 2048 point FFT at 44.1 kHz is mentioned as a typical value [12]. This corresponds to a frequency resolution of about 21.5 Hz. Taking the FFT of each individual frame of 160 points at 16 kHz will lead to a frequency resolution of only 100 Hz. To approach a value close to that of the original paper, the samples of multiple consecutive frames are combined. At 16 kHz a minimum of 745 samples per FFT is required to achieve a 21.5 Hz resolution. The nearest power of two above this number is 1024, powers of two are the norm for FFT operations. Thus the samples of the six most recent frames can be stored in a buffer, resulting in 960 samples. The buffer is updated each 10 ms to include the newest frame and discard the oldest if it is full. Every 10 ms a FFT is done using the contents of the buffer, always zero padding the contents to the nearest above power of two. This technique ensures sufficient frequency resolution after 60 ms of consecutive use. There is a start up time in which the frequency resolution is insufficient.

3.5. Testing with simple signals

The output of the Ocean algorithm is a frequency-domain signal that is pitch shifted. Figure 3.1 shows the resulting frequency spectra of the Ocean algorithm using a filtered audio fragment of a LP voice as input. The shifted spectrum (in blue) is stretched towards the right with respect to the original spectrum (in orange). The amplitudes remain the same. This means that the content of the frequency bins is successfully copied to the corresponding frequency bins a factor k higher. This is as required.

The Ocean algorithm is also tested by applying the Ocean algorithm to a simple sinusoidal signal of 1000

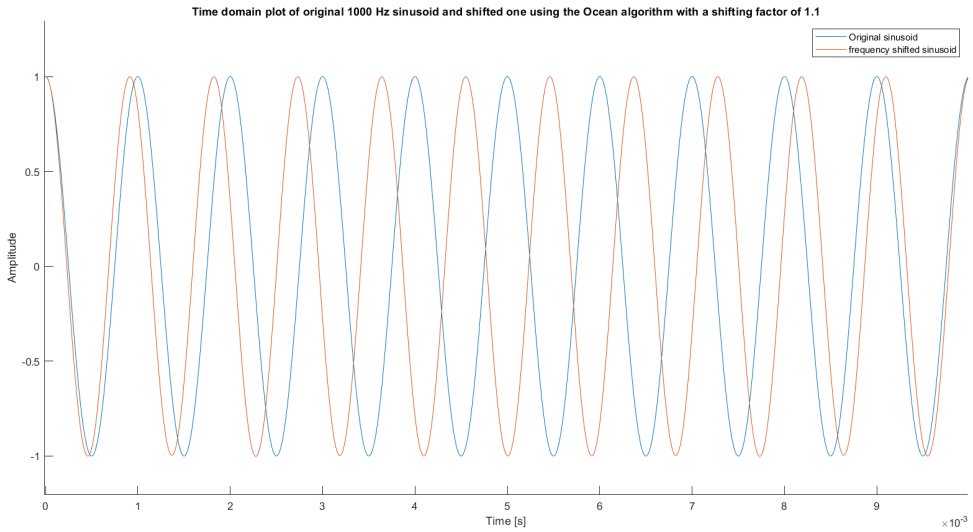


Figure 3.2: Test result of simple sinusoidal wave shifted with a shifting factor of 1.1

Hz to with a shifting factor of 1.1. If the result is transformed back to the time domain, it is clearly visible that the signal has changed to a 1100 Hz sinusoidal signal. When listening to the pitch shifted fragment without applying an overlap window, it contained a disturbing echo. After a couple of seconds the fragment started over again. Due to the large frame length without windowing and inaccuracy of the inverse fast Fourier transform (IFFT), the frequency components contain a great uncertainty in the time domain. When applying a Hann or Hamming window, the fragment is chopped into frames of $N = 160$ samples, i.e. 10 ms. This makes the echo inaudible. However, the result sounds very tinny. This can be adjusted by tuning the shifting factor, this will be done in Chapter 7 by measuring the intelligibility.

4 | Hardware

This chapter describes the chosen hardware components of the system. As mentioned in Chapter 2, the system consists of a data processor that is able to do the pitch shifting, an actuator that converts the modified voice into sound and an amplifier that drives the actuator.

4.1. Digital Signal Processor

Because the processing of data includes complex operations on the signal in frequency domain, a digital signal processor is chosen. The DSP has to able to run the Ocean algorithm described in Chapter 3. This should be possible since a DSP is a microprocessor chip which is designated for audio signal processing. It can take an analogue signal as input and convert it with a an analogue-to-digital converter to a digital signal. The digital signal can be processed by the DSP with operations such as the Fast Fourier Transform (FFT) which is very convenient for implementing filters and the pitch shifting. The output can either be an analogue and digital signal. Examples of digital output are I^2S and Pulse-Width-Modulation (PWM).

Specification

Digital signal processors are produced for a wide variety of applications. This design requires a DSP that is small and power efficient. It should also be suitable for audio applications and be able to perform operations on audio signals in real-time. In a later stage of design it is desirable to have a convenient testing environment. A software suite and demonstrator board to accommodate this need are well suited. There are also some hard demands the DSP has to meet in order to be implemented in the design.

In- and output

Its desired input must be compatible with the output of the microphone. This is an analogue signal with values between -0.3 V and $V_{DD} + 0.3$ V. The DSP uses digital signals, so an ADC must either be integrated or externally connected. The ADC must be able to handle the bandwidth up to 8 kHz, and thus have a sample rate of at least 16 kHz. The output of the DSP must be compatible with the input of the amplifier that drives the speaker. In the case of digital input amplifier, the protocols must match. In the case of an analogue input amplifier, a DAC must be either integrated in the DSP or externally connected.

Memory

Furthermore, the code of the different algorithms must be stored on the device, meaning there must be enough ROM available. While developing the code in MATLAB the combined size of the main file and all called functions would not exceed 10 kB. The data that is being processed will also be stored on the RAM of the DSP. There is an input and output filter, each containing about 10 ms of data. At a sample rate of 16 kHz this will be 160 samples, each containing 16 bits of information.

$$16b \cdot 441 = 2560b = 0.32kB \tag{4.1}$$

Equation 4.1 shows that the memory required for the storage of the audio data is very limited. The RAM will also be used for all the operations on the data. Computing the FFT will be the most demanding operation and will be most demanding in RAM. The memory usage depends on the size of the FFT. According the Roscoe and Burt the required data in bytes for performing a FFT, excluding buffers, is given by Equation 4.2 [28] for 32-bit arithmetic.

$$M_{required} = 8 \cdot N_{FFT} B \tag{4.2}$$

As described in Section 3.4 the FFT size will be 1024. This means 8196 bytes or 8 kB of memory must be allocated for the FFT operation.

Power consumption

The power consumption is desired to be less than 0.375 W, this has several reasons. For a higher power consumption a larger battery capacity is needed in order to maintain the 2 hours speaking time. A larger battery capacity will result in a larger battery in terms of size, which is undesirable. Another reason for a maximum power consumption is heat dissipation and since the device will be worn close to the body this will be uncomfortable for the user.

Selection

There are multiple companies that each offer suitable DSPs, they are ranked in a scoring system based on the aforementioned specifications. The scoring system can be found in Appendix B.1. The recommended devices of Texas Instruments (TI) and Analog Devices (AD) for audio applications are considered. These are devices from the C55xx family of TI, two devices from the Blackfin line and one from the Sharc line of AD. The criteria mentioned above (input voltage range, ROM, RAM, ability to perform FFT, power consumption) as well as the ease of programming and price of the device are combined in a scoring system. The criteria are weighted to indicate their relative importance. The DSP must be able to perform the algorithms, and so the amount of memory and ability to perform a FFT are rated as most important. The power consumption cannot exceed the budgeted 375 mW, if it does the device cannot be implemented in the design. Other criteria are of lesser importance, but are quality of life factors. The ease of implementation depends on how easily the input and output are connected to the rest of the system. Ideally there are an ADC and DAC present on the device itself. The ease of programming is dependent on the software development environment that accompanies the DSP. For the TI this is Code Composer Studio, for AD this is VisualDSP++.
The chosen DSP is the TMS320C5515. It comes with several features that are very relevant for the design, such as a dedicated FFT hardware acceleration block and an integrated ADC with a sufficiently high sample rate. The FFT block operates using 16-bit arithmetic. With 128 kB of ROM and 320 kB of RAM there will be no shortage of available memory. There is no integrated DAC, but there is an I^2S bus that makes integration with audio peripherals manageable [35].
Texas Instruments works in close cooperation with the TU Delft which provides ease of access to samples and support. In addition, the Code Composer Studio (CCS) is a free software suite available on the Texas Instruments website and will be suitable for testing the design later on.

4.2. Actuator

The output of the amplifier will be converted to sound by an actuator. Different available actuators are elaborated.

Dynamic speaker

In a dynamic or moving coil loudspeaker, an electrical signal is sent from the source through the coil. Due to the induced magnetic field en the field of the surrounding magnet, the coil moves. A cone is attached to the coil, which creates air pressure when moved. The amount of movement, and thus the level of sound production, depends on the electrical signal from the source. This type of loudspeaker is the most common one in use [25].
Similar to ribbon microphones, there are also ribbon speakers that use a thin strip of foil as both the conductor and the moving membrane. This type of design is mostly suited for higher frequencies [22].

Electrostatic speaker

In this design, a diaphragm is stretched between two perforated metal sheets called stators. When operating, a positive charge is applied to the diaphragm. The stators are connected to the output of the audio amplifier. The stators become positively or negatively charged based on the source signal. The positive charge on the diaphragm is repelled by the positive stator and attracted to the negative one. This causes movement and thus also sound production. Although this design can achieve very accurate sound production due to its light diaphragm, it is very difficult to achieve efficiency and output capabilities that rival the conventional moving coil design [19].

Piezo speaker

Operating on the piezo-electric effect, the diaphragm of this speaker moves when the piezo element attached to it changes shape due to an electric current. An advantage of piezo speakers is their form factor. They are much more compact than the traditional dynamic speaker while still delivering a large amount of sound production. They are also more power efficient, especially in the mid-range frequencies. A disadvantage is that piezo speakers require higher voltages to be driven. Piezo speakers also act as a capacitive load which requires specific amplifiers. Their frequency response is typically inferior to dynamic speakers, ceramic speakers can have a loud output at the cost of sound quality [13] [14] [29].

Specifications

In order to choose the right speaker for the system, some specifications for the speaker has to be prioritised. Most priority is given in the selection system of 4.2 to the frequency response of the speaker from the range between 2 kHz and 4 kHz. This is the frequency range which is most important for intelligibility of the speaking person [30]. Since this is the main goal of the system, this specification is given most priority. The criterion is based on whether the response high enough in this band and also whether it is 'flat', which means that the response is about the same over this range and there are no peaks and downs.
The minimum sound pressure level is also an important measure for the intelligibility. For this choice, the pressure at 0.3 m with 1 W applied is considered in decibels.
Although the frequency range of 2-4 kHz has highest priority, the human voice has a larger range. The minimum frequency is 500 Hz and the maximum is around 4 kHz [30]. In order to make the voice as natural as possible, the speakers output should match the total frequency range of the the human voice.
The rated power of the speaker is important for multiple aspects. With a higher rated power, the battery should have a larger capacity, which leads to either a larger battery size and therefore an increased total system size or the speaking time will be lower and the system needs to be recharged more often. A higher rated power can also cause more heat dissipation, which is not comfortable for the user of the system, since it will be worn close to the body.
The desire is to keep the size as small as possible. Although the product is in its very early stages and size is not of the utmost importance, a smaller option is preferred when performance in other categories is similar.

The last criterion for the speaker is the price of the speaker itself. Since it will not affect the performance of the system, the costs of the speaker is not prioritised very high, but it is taken into account with low priority.

Selection

When considering the available technology, there are two viable options: the conventional moving coil, and the piezo technology. They are both viable because they meet the minimum requirements but each excels in different areas. The moving coil will provide better frequency characteristics and thus better sound quality. However, the piezo speakers are much more compact and efficient. A number of viable speaker options is selected and rated using a score system as seen in Appendix B.2. Prioritised specifications got a bigger weight factor. As a result, the RS Pro from RS Components (RS stock code: 1170644) was selected as the best moving coil option. The SCS-32 is the best piezo option that is readily available.

4.3. Amplifier

Since the DSP is not capable of driving the speaker, an amplifier is needed to do so. Amplifiers can be subdivided into different classes. Furthermore, a distinction is made based on the type of input signal the amplifier takes, analogue or digital.

Classes

There are different classes of amplifiers, the most prevalent in audio applications being A, B, AB and D as seen in [10]. Class A uses a continuously biased transistor, class B uses two switching MOSFETS. Class AB, as its name implies, is a crossover between class A and B. Class D amplifiers is a switching amplifier and uses Pulse Width Modulation (PWM). The most notable difference between the classes is the efficiency (and therefore heat dissipation), the class D amplifier being the most efficient with up to 90% efficiency. Classes A, B and A/B lag behind at 60% or below. Class D amplifiers do have bandwidth limitations due to its low-pass filter and converter module.

Input

The output of a DSP can be analogue and/or digital. This has a great impact on the selection of the amplifier. An analogue input amplifier is very straight forward to use: just increase the amplitude. However, a digital input amplifier could produce less noise and could save a DAC and switching circuitry when using a class D amplifier. The data has to be transferred using a communication protocol such as I^2S and could cause extra delays.

Specifications

The amplifier must amplify the signal received from the DSP and drive the selected speaker such that there is enough sound production with adequate quality. This means that it must be able to deliver at least 1 W

of power to an 8 Ω load, preferably up to 1.5 W. To retain the quality of the input signal, the SNR and total harmonic distortion (THD) must be kept as low as possible.
As is the case with all components in the system, efficiency is very important. Higher efficiency means that less capacity and thus less space is needed for the power source. Additionally, power dissipation can lead to uncomfortable heat production which is especially important in a wearable device.
The volume control of the system will be included in this component. A built-in volume control in the off-the-shelf component is a nice to have, but not a must. Lastly, the price of the component is taken into consideration.

Selection

With the above specifications in mind, a comparison between some readily available components is made. Using a scoring system, each amplifier is awarded points for its performance in each specification criterion. The selection criteria have a weight factor based on their importance. The highest being the ability to deliver enough power to the speaker. Efficiency comes second. Because efficiency is so important it is an obvious choice to only look at class D amplifiers or amplifiers with a digital input.

Resistive load
The amplifiers that are suited for dynamic speakers are compared. The scores can be found in Appendix B.3. The LM4675 is found as the best option with an analogue input. For a digital input, the SSM2529 is determined to be the best option.

Piezo
The data sheets of the amplifiers suited for piezo speakers do not offer information on all of the selection criteria used for the previous scoring system. They are compared in a different table using metrics that are defined for all components. The scores can be found in Appendix B.4. All the components in this table, with the exception of the TPA2100P1, are listed as recommended pairings with the SCS-32 speaker that was found as the most suitable piezo speaker for the design. The TPA2100P1 is selected, its efficiency being its largest selling point. It is the only class D amplifier on the list.

5 | Additional features and expected performance

5.1. DSP

Low power mode

The DSP cannot be disconnected from the power source completely and still operate correctly. However, when the user is not speaking the DSP can enter a low-power mode to improve efficiency.
The device can enter an idle mode using the Idle Control Register (ICR). The power consumption in standby is given by TI as 0.7 mW. The device can be brought out of standby mode using the WAKEUP pin. If a pulse of longer than 30μs is sent to this pin, the device wakes up and continues regular operation. This is a very useful feature for a design that is only meant to be used for a few hours a day. The pulse at the WAKEUP pin will be provided by the air pressure sensor [35].
In the power budget drawn up in the other subsystem, 0.375 W is set as the maximum power consumption for the DSP. This number was chosen assuming that no power would be consumed when the user is not speaking. Due to the consumption in standby mode and the consumption of passive elements in the surrounding system which will be described in Chapter 6, the passive power consumption is at least 1.12 mW. The budgeted 0.375 W for two hours correspond to 2700 J of energy. If the device is worn for 12 hours without charging, this means it is in low power mode for 10 hours. This corresponds to 40.32 J of energy spent on low power mode, leaving 2659.68 J for the two hours of speaking time. The maximum power consumption while speaking is thus adjusted to 0.3694 W.

Algorithm performance

This section describes the performance of the DSP when it is executing the algorithms using different clock speeds. The performance will be measured in delay and required energy for the operations. Because there will not be a possibility to test the specific algorithms that have been described in this thesis, the performance is gauged using documentation that TI has made available.
There is an application note on the FFT implementation on the C55x5 processors. In this document, the performance of the FFT Hardware Accelerator (HWA) is compared to that of the CPU of DSP using the amount of required cycles and energy to perform one FFT. The relative advantage of the HWA over the CPU depends on the length of the FFT, but not on the clock speed the device runs on. The tables can be found in Appendix D.4. The HWA can perform transforms of lengths 2^N with a minimum of 8 and a maximum of 1024. At the maximum length, the difference in performance is the largest. The HWA requires 3.8 times less cycles and 6 times less energy. In absolute terms the HWA requires 7315 cycles. At 60 MHz this is equivalent to 122μs and requires 1836.2 nJ of energy. At 100 MHz, the FFT takes 73.15μs but requires 2820.6 nJ. There is a clear trade off between power and speed. In the algorithms, three FFT operations are performed on each frame. Since each frame is 10 ms and there will be a frame overlap of 50%, 200 frames will pass through the algorithm every second. Thus 600 FFT operations are done per second. Even when using the more energy hungry 100 MHz mode, the power required for the FFT operations is $\text{FFT/s} \cdot \text{energy/FFT} = 600 \cdot 2820.6nJ = 1.7mW$. The delay in ms per frame is given by $\text{FFT/frame} \cdot \text{ms/FFT} = 3 \cdot 73.15 \cdot 10^{-3} = 0.22ms$. The delay for the slower 60 MHz mode is $\frac{100}{60} \cdot 0.22ms = 0.37ms$ [34]
The trade off between energy usage and operating speed is also seen in the typical power consumption estimation of the device using different clock speeds. A table is available in the Appendix. The estimated power consumption is based on modelled typical utilisation of the device. For 60 MHz, the estimated power consumption is 16.7 mW and for 100 MHz it is 51.9 mW. It appears that the calculation of the FFTs will only constitute a fraction of the total power consumption [37].
Unfortunately, there is no documentation that gives insight into the power consumption of the other operations in the proposed algorithms. However, as the FFTs are most likely the most demanding operations it is presumable that the power budget of 0.3694 W will not be exceeded. The delay per frame of 0.22 ms is also an encouragingly low number because it is relatively small compared to the 10 ms delay that is inherent to the system.

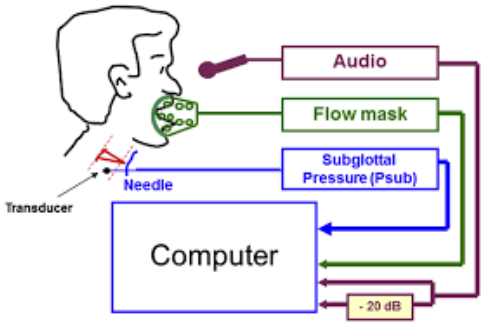


Figure 5.1: Measurement setup for subglottal pressure.

System control

The DSP will not only send the data to the amplifier, it will also regulate the volume. The DSP receives the signal of the pressure sensor and based on the readings will determine how loud the user wishes to speak. The method of volume control will differ per amplifier. Possible implementations include writing to registers in the amplifier through I2C interface, or sending a signal to a voltage controlled resistor in the amplifier network.

5.2. Pressure sensor

One of the extra desires of the system is hands-free volume control, this can be achieved with a pressure sensor in the trach. Since speech is based on air pressure, this can be used for volume detection of the desired speech. LP still use the same air pressure to speak as people with normal speech, however there is no larynx to modulate this air pressure to clear speech. This air pressure, the subglottal pressure, should be measurable with an air pressure sensor inside the throat. The trach is placed closely to the removed larynx and in direct contact with the subglottal pressure. The air pressure sensor could therefore be placed inside or attached to the trach, if the trach is fixed in place.

Range

The volume can be linearly proportional with the subglottal pressure if the air pressure sensor is accurate enough over the small range of the subglottal pressure. The subglottal pressure is in a range of 9.5 cm H2O[1] (± 2.8 cm H2O) for breathy phonation to 20.1 cm H2O (± 2.6 cm H2O) when the person is speaking loudly, this corresponds to 932 Pa to 1971 Pa respectively [31]. When silent the air pressure is around 100 Pa. However this can differ from person to person. The MPX5100 air pressure sensor [8] can for example be used for this purpose since it can be either used as a differential as well as an absolute air pressure sensor. The MPX5100 has a sensitivity of 45 mV/kPa. This means a voltage range of roughly 85 mV for speaking loudly and around 40 mV for normal speech. This is, however, a quite small range and could be improved with an air pressure sensor with a higher precision [1].

Output

The output of the air pressure sensor should be connected to the DSP with a pre-amplifier to match the voltage range of the air pressure with an appropriate input voltage range of the ADC of the DSP. The DSP should translate the measured value of the air pressure sensor to the digital input of the amplifier for the gain control. Figure 5.1 shows the setup for the subglottal pressure measurements done by Sundberg [31]. Instead of the needle, the air pressure sensor can be permanently used in the stach of LP. The computer is then replaced by the DSP.

The entire idea of the hands-free volume control is still a concept and further implementations are described in the recommendations.

[1] 1 cm H20 = 98.0665 Pa

6 | System integration

6.1. DSP peripheral circuit design

The pin map of the DSP is shown in Appendix D.1. Not all 196 pins will be in use for the current design and thus not all connections will be considered. However, all relevant pin connections will be described in this section.. An functional overview of the connections is given in Figure 6.1. The size of the DSP chip is 10.1 x 10.1 x 0.94 mm excluding all peripheral components.

Power

In order to use the device, there are some pins that must always be connected [38]. The CV_{DDRTC} pin must always be supplied between 0.998 V and 1.43 V; a typical value of 1.3 V is recommended. Considering the battery delivers 3 V, there must be some type of voltage drop to supply the pin correctly. This can be achieved using a divider consisting of resistors. Using a resistor of 15 kΩ and one of 6.5 kΩ and placing the pin in parallel with the smaller resistor leads to the correct voltage at the pin. Although this is not the most power efficient technique, it is very cheap to implement The passive dissipation is 0.42 mW, which does not greatly impact the total power consumption. The total budget for the DSP is set at 375 mW, as mentioned before. Using higher resistances, the power dissipation could be lowered. However, the maximum current rating of the pin does not allow for higher resistances. The DV_{DDIO} pins supply power to the I/O pins of the device. These pins must always be powered for proper operation. They can be connected to the power source directly. The DV_{DDRTC} powers the WAKEUP pins which will be discussed later. The supply pin must always be connected to the 3V battery.
The LDOI must also always be connected. The acceptable range is 1.8 V to 3.6 V and can thus be connected to the battery directly. The pins are the power supply for the internal Low Dropout Regulators (LDOs). These modules are used to regulate the power supply of other modules on the device. The ANA_LDO regulates for the analogue PLL and the ADC, the DSP_LDO for the digital core, and finally the USB_LDO for the USB core. For the current design, the digital core will be very important as this is where the calculations take place. The core voltage CV_{DD} can be supplied by the DSP_LDOO, which is the output of the DSP_LDO. For proper operation, a 5 μF - 10 μF external decoupling capacitor is recommended. To enable the DSP_LDO, the $\overline{DSP\_LDO\_EN}$ pin must be connected to ground. ANA_LDOO is the output that supplies the V_{DDA_PLL} and V_{DDA_ANA}. The recommended external decoupling capacitor is 1 μF.
Furthermore, it is important to not leave an of the input pins floating as they will consume power. There are internal pull-up and pull-down resistors that can manage the pins through software. The settings can be configured in the Pullup/Pulldown Inhibit Registers. When the DSP is not used as a USB device, the USB_LDOO must be left floating and disabled in the LDO control register, the USB_R1 must be connected to ground through a 10 kΩ resistor and 14 other USB pins are connected to ground directly.

Memory

Because the DSP was chosen such that the internal memory is sufficient to store the algorithms, there is no need for an external memory interface (EMIF). This greatly simplifies the circuit design, as the EMIF pins can be grounded. These 58 pins are listed in the TI manual in Table 2-8.

Control

The DSP receives a control signal from the air pressure sensor. Because this is an analogue signal, it will enter the DSP via the 10-bit SAR ADC that is on the device. This module is accessed using the general purpose analog input pins (GPAIN3:0). The signal is also connected to the WAKEUP pin to get the device out of standby mode when the user is speaking.
The DSP outputs a control signal to the amplifier to regulate the volume control. In the case of the SSM2529 this is done using the I2C interface (SCL and SCA pins). This implementation is shown in Figure 6.1. For the piezo implementation this will be one line that carries a signal to the gain select pin of the amplifier.

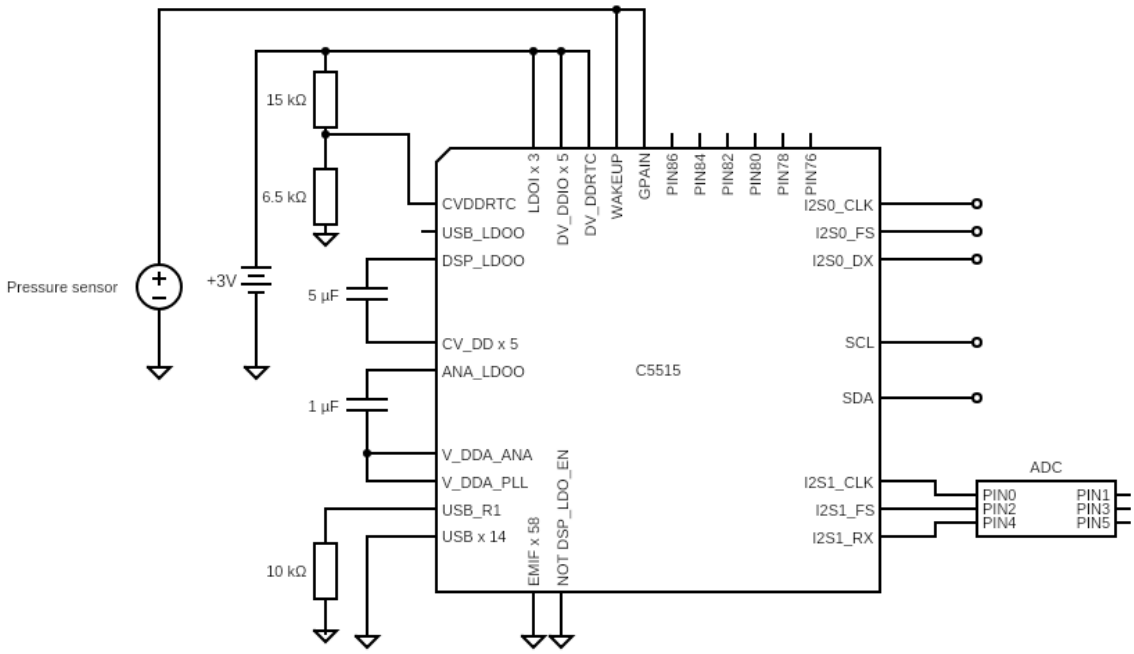


Figure 6.1: Functional overview of the DSP pins and their connections for the moving coil implementation

Data

The DSP uses two I^2S busses for data communication. One is used to receive the input signal from the ADC. The other is used as output towards the amplifier. If the piezo implementation is used, the DSP is connected to a DAC because the TPA2100P1 does not accept digital inputs. The SSM2529, however, does have an I^2S interface as is described in Section 6.3.

6.2. Amplifier and speaker circuit design

As described there are two implementations, namely the moving coil option and the piezo speaker option. Both options need different circuitry. Both amplifiers provide an option to control the volume and to put the amplifier into sleep mode/low-power mode. The volume control could be implemented using a pressure sensor near the LP throat as described in Section 5.2. Due to time constraints the volume control and power-down control have not been implemented. A recommended implementation is described in Chapter 10.

Option 1: Moving coil

The circuit design for the moving coil speaker and accompanying amplifier can be seen in Figure 6.2. This block diagram is taken from the datasheet of the SSM2529 digital amplifier [4] in hardware mode. The audio processor corresponds to the DSP. The DSP and SSM2529 communicate throughout the I^2S protocol and will be described in Section 6.3. The System Controller is a system that could control the volume and put the amplifier into sleep mode/low-power mode via I^2C communication busses.
The remaining pins of the SSM2529 are described in Table 6.1. They are used mainly to connect the speaker and the power supply. The SSM2529 is set in hardware mode to connect with other devices. The component measures 1.96 x 1.98 x 0.56 mm.

Table 6.1: Remaining pins of the SSM2529

Pin name	Description	Value
IOVDD	I^2C bus power supply	3V
LDO_OUT	LDO output	Not in use
DVDD	Digital power	1.3V
SPKVDD	Speaker power	3V
SPKGND	Amplifier ground	0V
GND	Digital and analogue ground	0V
SA_MODE	Standalone or Hardware Selection	0V (Hardware Mode)
ADDR	Interface Select in Standalone Mode	0V (Hardware Mode)

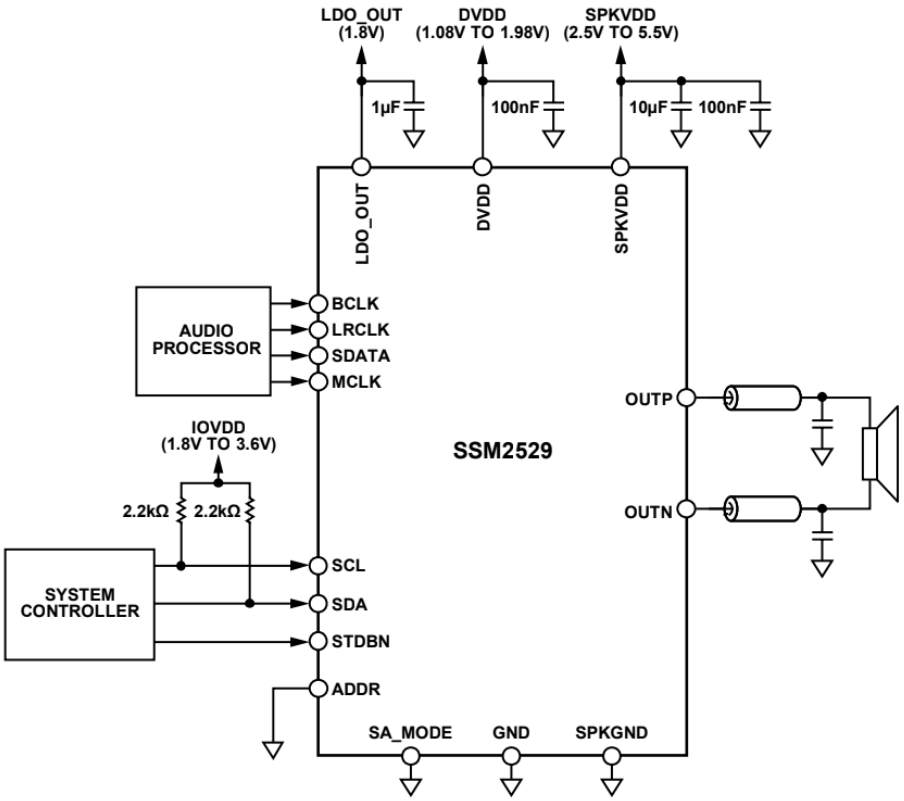


Figure 6.2: Functional block diagram of the moving coil speaker and the accompanying digital amplifier connections.

Option 2: Piezo

The circuit design for the piezo speaker and accompanying amplifier can be seen in Figure 6.3. This block diagram is taken from the datasheet of the TPA2100P1 analogue piezo amplifier [33]. The Analog Baseband or CODEC corresponds to the DSP and provides the input. The Digital Baseband is able to control the volume and put the amplifier into sleep mode/low-power mode. The dimensions of the device are 2.3 x 2.2 x 0.625 mm.
The VIN/Vdd pin is connected to the supply voltage of 3V. The VccOUT and VccIN pins are connected. The VccOUT pin is the output of the build-in boost converter, which boosts the voltage to a max of 9.5V/19V_{pp}.

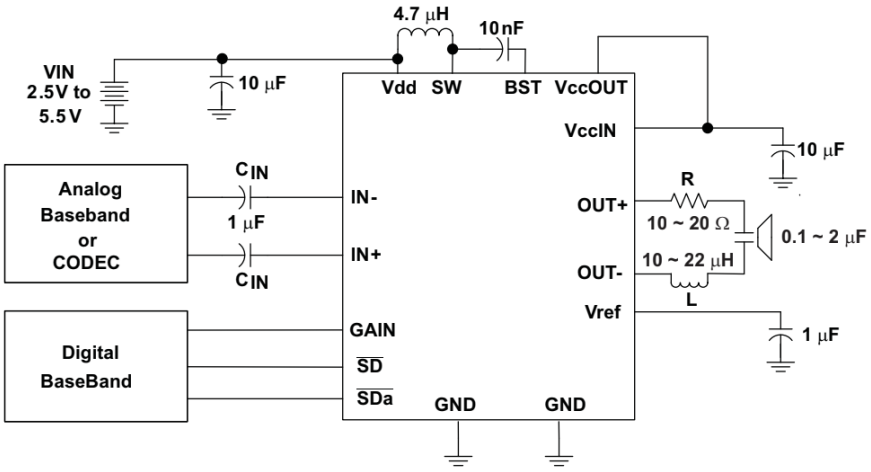


Figure 6.3: Functional block diagram of the piezo speaker and the accompanying analogue amplifier connections.

The value of R and L in series with the speaker can be determined using Equation 6.1 and Equation 6.2 respectively. The highest frequency f_{max} is 8 kHz. The nominal speaker capacitance C_{spk} is 66 nF with a 20% deviation. When using Equation 6.1, C_{spk} is set to its highest possible value of 79.2 nF to determine the minimal resistance of R. This results in $R_{min} = 250\ \Omega$.
When using Equation 6.2, it is desired to make the output load configuration overdamped ($\zeta > 0.707$). When ζ is greater than one, then the maximum audio frequency will be limited by the resistor – speaker capacitance low pass filter as shown in Equation 6.1. Therefore ζ is set to 1 and C_{spk} is set to its lowest possible value of 52.8 nF to determine the maximal inductance of L. This results in $L_{max} = 820\ \mu$ H for R = 250 Ω.

$$R = \frac{1}{2\pi f_{max} C_{spk}} \tag{6.1}$$

$$L = \frac{R^2 C_{spk}}{4\zeta} \tag{6.2}$$

Boost converter
The TPA2100P1 has its own boost converter to deliver a high peak-to-peak voltage necessary for piezo speakers. The boost converter can be tuned via the inductor and capacitor to the SW pin. The boost inductor stores current, and the boost capacitor stores charge. As seen in the block diagram (6.3), typical values are 4.7 μH and 10 nF respectively. However, these values are actually determined by the requirements of the speaker.
The inductor selection is done using Equations 6.3, 6.4 and 6.5. The load voltage is set to the max peak-to-peak voltage the amplifier can deliver. This is 19 V_{pp}, and therefore V_{CC} = 9.5 V. The switching frequency of the boost converter f_{BST} has a typical value of 1.2 MHz.

$$I_{CC} = I_{max} = \frac{V_{CC}}{|Z_{min}|} = V_{CC} \cdot 2\pi f_{max} C_{spk} = \frac{V_{CC}}{R_{min}} \tag{6.3}$$

$$I_L = I_{CC} \frac{V_{CC}}{V_{DD} \cdot 0.8} \tag{6.4}$$

$$L_{BST} = \frac{V_{DD}(V_{CC} - V_{DD})}{I_L \cdot f_{BST} \cdot V_{CC}} \tag{6.5}$$

This results in a value of $L_{BST} = 13.71\ \mu$H.

The capacitor value can be selected using Equation 6.6. Best practice is to use low ESR capacitors to minimize the ripple on the output voltage. Ceramic capacitors are a good choice if the dielectric material is X5R or better. When selecting a ceramic capacitor, the capacitance has to be chosen twice as high as theoretical calculated since the working capacitance drops with DC drastically, compared to aluminium or tantalum capacitors. Therefore k should be at least 2 for ceramic capacitors, and the rated voltage should also be at least set twice as high as needed.

$$C_{BST} = k\frac{I_{CC}(V_{CC} - V_{DD})}{\Delta V \cdot f_{BST} \cdot V_{CC}} \tag{6.6}$$

The desired ripple voltage ΔV is set to 100 mV_{pp} since audio is very sensitive for noise. For a ceramic capacitor, its capacitance should be at least 359.3 nF.

6.3. DSP-Amplifier interface

I2S

The Inter-IC Sound (I2S) interface is commonly used in device to device digital audio transfer. The I^2S bus consists of at least 3 data lines between the master (DSP) and slave (amplifier) device. The DSP has four of these module which can be individually configured. The first line is the bit clock (BLCK) which defines the length of one bit in the signal. The second line is the left-right clock (LRCLK) or word clock (WCLK). A 0 on the LRCLK corresponds to the left audio channel, a 1 with the right channel. The final line carries the data signal (SDATA). The interface is illustrated in Figure 6.4. This figure is taken from the manual of the DSPs I^2S module, which uses a slightly different naming convention. The following section relies heavily on the information provided by the TI I2S manual [36]. I2S_CLK is the bit clock, I2S_FS is the left-right clock and I2S_DX is the data output line. The I2S_RX line is used for incoming data, but this will not be used for the interface between the DSP and the amplifier.
The I^2S module of the DSP supports two data communication formats: the general I^2S/left-justified format and the specialised DSP format. Because the DSP mode is the only one that supports mono sound, this format is chosen. The timings of this format are shown in Appendix D.2. The key difference between this format and the more standard I^2S/left-justified format is that the frame synchronisation uses pulses, instead of being high for left and low for right.

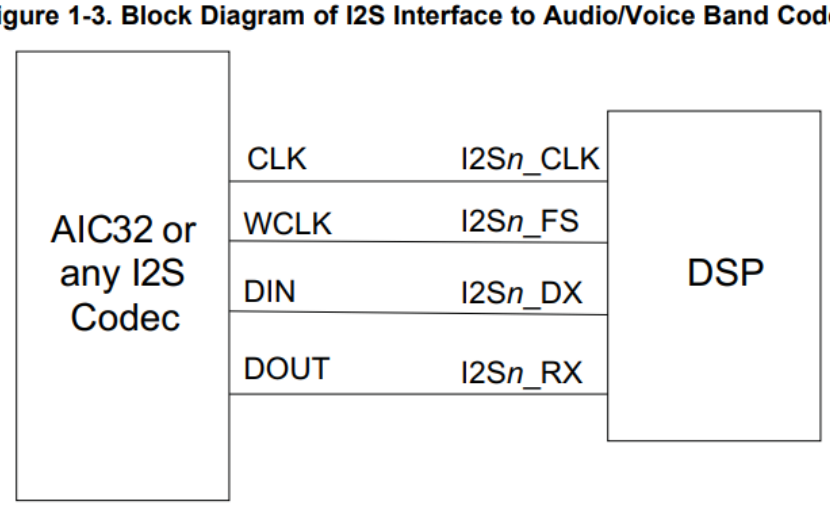


Figure 6.4: Illustration of I2S interface taken from TI user manual [36]

Configuring the DSP

The DSP is configured in the master mode. As the master device, the DSP will determine the clock rates. The I^2S clock rates can be set according to Equations 6.7 and 6.8. They are set to fractions of the system clock.

$$I2Sn\_CLK = \frac{SystemClock}{2^{\text{CLKDIV+1}}} \tag{6.7}$$

$$I2Sn\_FS = \frac{I2Sn\_CLK}{2^{\text{FSDIV+3}}} \tag{6.8}$$

The standard audio format used in the device is 16 kHz with 16 bits per sample. To achieve this level, the word length will be set to 16 bits. If the sample rate is to reach 16 kHz, the bit rate must be 16 times as high:

Table 6.2: I2SSCTRL configuration (DSP)

Bit	Name	Value	Description
15	Enable	1	The module is enabled
14-13	Reserved	-	-
12	MONO	1	Mono mode is selected
11	LOOPBACK	0	Loopback is disabled
10	FSPOL	1	Frame synchronisation is pulsed high
9	CLKPOL	1	BCLK rising edge is used to sample
8	DATADLY	0	Data delayed by one bit
7	PACK	0	Data packing mode is disabled
6	SIGN_EXT	0	No sign extension
5-2	WDLGNTH	0100	16-bit data word
1	MODE	1	Master mode
0	FRMT	1	DSP format

I2S_CLK = 256 kHz. The bit clock is a fraction of the system clock. Choosing a value for CLKDIV determines the required system clock. For example, CLKDIV = 5 will correspond to a system clock of 8.192 MHz while CLKDIV = 6 corresponds to a system clock of 16.384 MHz. In Section 5.1 it became apparent that a 60 MHz clock will be more than sufficient for performance with a delay of 0.37 ms per frame due to FFT calculations. Because it is unknown what delays the other operations will cause, it is wise to not limit the system clock speed too much based only on the FFT performance. Choosing CLKDIV = 8 results in a required system clock of 65.536 MHz. Using the PLL module of the DSP, a system clock can be selected as multiple of the reference clock of 32.768 kHz as described in Equation 6.9. In this equation, M is the register in which the multiplier is stored.

$$PLL\_OUT = 32.768kHz \cdot (M+4) \tag{6.9}$$

By choosing M = 1996, PLL_OUT becomes 65.536 MHz. A schematic overview of the clock generator is shown in Appendix D.3. To put the DSP in PLL mode, the CLKSEL pin must be connected to a logical 0 (ground), the RDBYPASS bit must be set to 1 as there is no need for a reference divider. The PLL_OUT also does not need to be divided, so OUTDIVEN must be set to 0. Finally SYSCLKSEL mest be set to 1 to select the PLL_OUT as the system clock.
The settings for the I^2S module on the DSP are configured in the I2S Serializer Control Register (I2SSCTRL). The chosen values for each of the settings in this control register are denoted in Table 6.2.

Configuring the amplifier

The amplifier is configured in slave mode. This is done by connecting the stand alone pin (SA_MODE) and address pin (ADDR) to ground. It receives its clock from the master device: the DSP. The LRCLK can be used to generate the internal clock of the amplifier. The required internal clock speed is dependent on the received sample frequency, in this case this frequency is 16 kHz. In the data sheet, multiple settings are recommended for this sample frequency in Table 48. The MCLK is chosen as an integer multiple of the sample frequency. Using a ratio of 512, the MCLK is set at 8.192 MHz which is within the range of recommended clock speeds. In the block diagram in Figure 6.5, it is shown how the internal clock is generated. In the digital PLL the LRCLK, which runs at the sample frequency, is multiplied by 2^N. By taking N = 9, the right clock speed achieved. The clock must not be changed by the analog PLL. The relationship between the input and output frequency of the APLL is given in Equation 6.10 [4].

$$f_{PLL} = f_{in} \cdot \frac{R + \frac{N}{M}}{X} \tag{6.10}$$

R, N, M and X are values defined in the PLL registers. The factor in Equation 6.10 must amount to 1. Because N = 9, this can be achieved by choosing M = 9, R = 0 and X = 1. The settings for the serial audio interface of the amplifier are determined by two registers, SAI_FMT1 and SAI_FMT2. These registers are configured to be compatible with the signal the DSP provides. The chosen values are denoted in Tables 6.3 and 6.4.

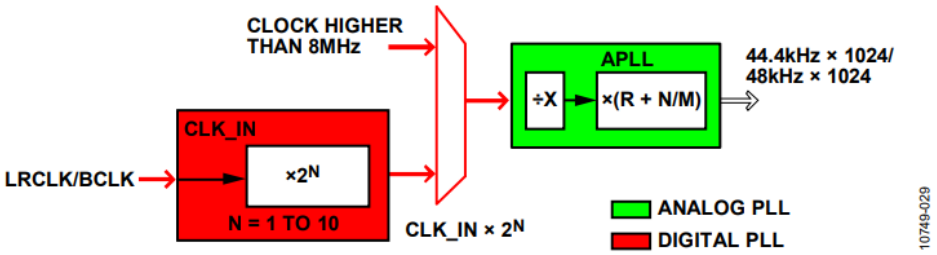


Figure 25. Clock Generation Block Diagram

Figure 6.5: Block diagram describing the generation of the internal clock of the amplifier as taken from the data sheet [4]

Table 6.3: SAI_FMT1 configuration (Amplifier)

Bit	Name	Value	Description
7-6	SDATA_FMT	00	I2S, BCLK delay by 1
5-3	SAI	101	Mono PCM
2-0	SR	101	Sample rate: 16 kHz

Table 6.4: SAI_FMT2 configuration (Amplifier)

Bit	Name	Value	Description
7	LPST	0	Small power stage disabled
6-5	LR_SEL	x	Don't care
4	LRCLK_MODE	1	Pulse mode LRCLK
3	LRCLK_POL	0	Normal LRCLK operation
2	SAI_MSB	0	MSB first SDATA
1	BCLK_TDMC	x	Don't care
0	BCLK_EDGE	0	BCLK rising edge used to sample

7 | Simulations

7.1. Pitch shifting

The pitch shifting principle was successfully implemented in MATLAB as described in Chapter 3. However, the shifting algorithm is also tested with test signals. The test signals that are used are short audio fragments form a laryngectomised person before and after his operation. The audio fragments contain the person speaking vowels, such as 'aaaa' and 'oooo', but also Dutch sentences which are used very often in speech analysis. The sample frequency of the test signals is 44.1 kHz. This is resampled to 16 kHz, since this is the sample frequency of the total device. Resampling is done by applying a FIR lowpass-filter to the signal. The difference between the two sample frequency is hardly audible.

Complications

The algorithm has thus far only been tested on a simple sinusoid, without dividing the signal into small frames on which to operate. To be able to play the shifted signal in real time, it will be necessary to divide the signal in to short frames to keep the delay below 15 ms.
Using the same test signal as described in Chapter 3, the algorithm is now applied to overlapping frames which are then buffered and added to produce a result. When functioning properly, this should reproduce the sinusoid at a higher frequency and resemble the result shown in Figure 3.2. The actual result of the test is shown in Figure 7.1.

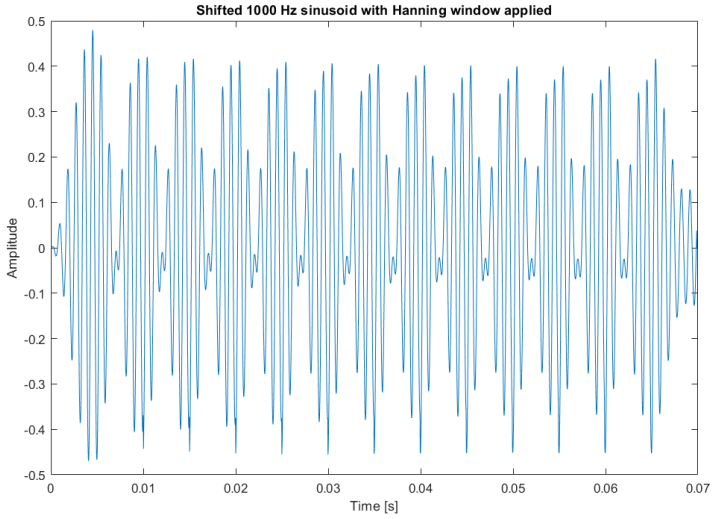


Figure 7.1: Ocean algorithm applied on simple sinusoid with windowing.

Although the frequency is increased, there are now points in the signal where the amplitude dips. These dips become wider as the shifting factor k is increased. When a speech signal is used as input, these complications are not apparent. The output signal remains the same length, does not have notable variations in loudness and is pitch shifted successfully. Testing using speech signals can be resumed.

Testing intelligibility with STOI algorithm

The factor k needs to be chosen such that there is a good trade-off between intelligibility and whether it sound naturally, which sounds most pleasant for the listener. For determining the intelligibility of an audio signal, several algorithms can be used. The Short-Time Objective Intelligibility (STOI) measure can be used to measure the intelligibility of an audio signal by comparing it with a clean unprocessed signal [2].
In this particular case the clean unprocessed speech is a recording of the same patient before the laryngectomy. It is not expected that the processed speech will exceed this level of intelligibility. The value at k = 1

represents the relative intelligibility of the voice as it sounds without shifting the pitch. The pre-operation and post-operation fragments are modified to be time aligned as precisely as possible, because this is one of the requirements for the STOI algorithm to work. The fragment recorded after the operation is shifted with 100 different shifting factors and compared with the fragment from before the operation. The result is shown in Figure 7.2. The values on the vertical axis does not have an absolute meaning, but it can be used to compare the fragments relatively to each other in terms of intelligibility.

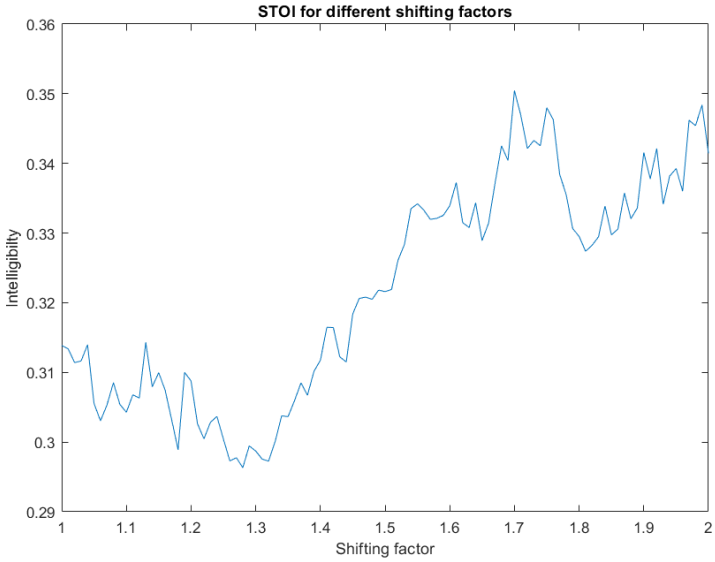


Figure 7.2: Intelligibility outcome of STOI measure for different shifting factors.

As can be seen in the figure, there is a dip of the intelligibility at around 1.3 and then rises again until 1.7. However, with shifting factors higher than 1.3 the pitch of the of voice is very high and does not sound naturally anymore, which is an important design desire. To make the make a good trade-off between intelligibility and naturalness, a shifting factor of 1.13 is chosen, since this is the most intelligible for shift factors that sounds natural.

Testing intelligibility with SIIB algorithm

Another method for testing the intelligibility is the Speech Intelligibility In Bits(SIIB) algorithm [42]. This algorithm is quite similar to the STOI algorithm, but a less precise time alignment is needed. The output of the algorithm is analogous to the STOI algorithm, the result will only be a value which can be compared relatively to another result. The same procedure as for the STOI algorithm is used for the SIIB algorithm.

At first glance, their seems to be no clear correlation to draw conclusions from it. However, the shifting factor of 1.13 provide again the highest intelligibility outcome for the low frequencies and therefore verifies the intelligibility outcome of the STOI algorithm.

7.2. Amplifier

Both amplifiers are attempted to implement in Simulink. However, the data sheets did not provide enough information to complete the Simulink models. As seen in Appendix E.1, the SSM2529 digital moving coil amplifier consists of several functional blocks but the functionalities of these blocks are not elaborated within the data sheet. The TPA2100P1 piezo amplifier data sheet did not even include a functional block diagram. For this reason, Simulink could only be used to create a general class D amplifier model which consists of a sawtooth generator, comparator and a filter/DAC at the output as seen in Appendix E.2. This resulted in an ideal working amplifier model but did not give valuable information about the gain, noise, distortion, frequency response, or any parameter of that kind. The only thing that could be simulated is the output filter/DAC which will now be described. Moreover, this includes that the simulation results will be not decisive in the choice between the moving coil and piezo speaker.

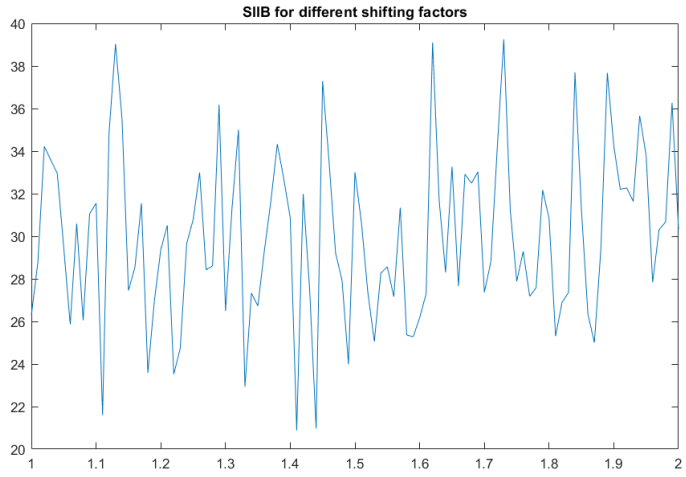


Figure 7.3: Intelligibility outcome of SIIB measure for different shifting factors.

The output filter consist of LC high-pass/DC filter, consisting of an 33 μ H inductor and an 680 nF capacitor. This is cascaded with an low-pass filter, consisting of an 1 mF capacitor and the speaker load. In case of the moving coil speaker, the load equals 8 Ω. This gives a flat response up to 4 kHz but blocks DC to protect the speaker, as seen in Appendix E.3. For higher frequencies, the output filter response goes to zero which filters out the frequency of 300 kHz of the sawtooth generator as seen in Appendix E.4. This results in an analogue signal which can drive the speaker. This can be seen in Figure 7.4. The blue signal is an LP audio fragment and its amplified version is visible in yellow. When opting for a different moving coil speaker in the future without a DAC included, an output filter as described could be implemented.
A piezo speaker amplifier, however, cannot be simulated in this way since the load is capacitive and the values for the compensation are not described by the TPA2100P1 data sheet as mentioned above.

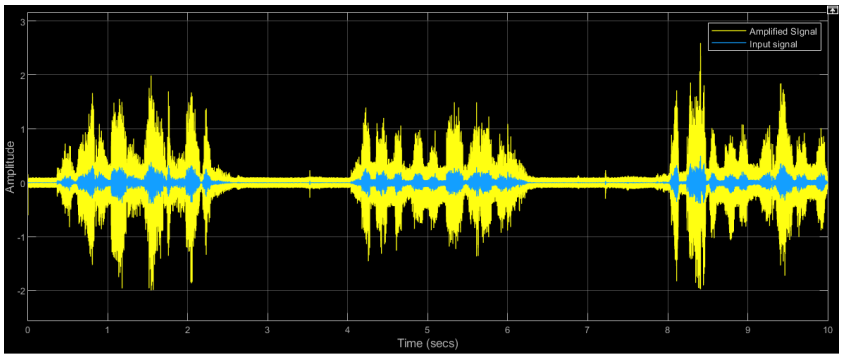


Figure 7.4: Simulation results for the amplifier in Simulink, with LP audio fragment.

8 | Discussion

8.1. Results

Shifting

The goal of shifting the speech signal is to improve the speakers' intelligibility and natural sound. The method used in this thesis is the Ocean algorithm. This algorithm proved successful at pitch shifting, first for simple sinusoidal signals and later for the patient's voice. There are measures for intelligibility using the two testing algorithms. The sweep with k ranging from 1 to 2 shows many peaks for both tests. For each maximum the shifted voice fragment was played to judge whether it sounds natural. This led to discarding all shifting factors above 1.3 due to the highly unnatural sounding squeaky voice. Both algorithms have a coinciding peak at k = 1.13 for optimal intelligibility while still sounding natural. The improvement of intelligibility is measured objectively using algorithms but the judgement of the naturalness of the resulting fragments is very subjective. Such subjective would be better left to a test panel whose scores would determine the most pleasing sounding shifting factor.
The complications described before could also have an effect on the intelligibility, even the effects are not directly recognisable. It is not known what causes the issues when the algorithm is tested using frames. It seems as though the length of the individual frames is decreased as the frequency is increased and that therefore the addition of the overlapping frames does not add up to the original amplitude at all points. However, this effect does not appear when the algorithm works on a single frame of a large size. It is also possible that the phase correction is off and that there is elimination due to the addition of out of phase signals. In other research, the method of phase correction proposed by the authors of the Ocean algorithm has been criticised [26]. But this does not explain why the dips in amplitude become wider as the frequency is shifted more.
The expected performance of the DSP in terms of delay is very promising for the Ocean algorithm. This means the current technology might be able to handle more intensive algorithms that can possible achieve better results for intelligibility. Further recommendations on this topic are given in Chapter 10.

Amplification and speaker

The goal of the sound actuation is to produce the improved voice with a SPL of 85dB at 0.3m. The size of the speaker plays an important role in the dimensions of the voice amplifier. Therefore a size versus sound quality and sound pressure trade-off has to be made. For this reason, a moving coil speaker and piezo speaker with accompanying amplifier have been pre-selected by comparing data sheets. Unfortunately, these components were not physically tested. Simulations were done instead (see Section 7.2) but did not provide decisive information. For this reason no specific speaker with amplifier is selected. Future research and physical testing have to be done to select a speaker and appropriate amplifier. Recommendations are described in Chapter 10.

8.2. Meeting the requirements

General requirements

The different parts of the system have been designed to meet the general requirements as well as possible. A data processing unit, amplifier and speaker have been proposed and the integration of all parts has been described. Although no exact size limitations were discussed at this early point in development, the final product should be small enough to wear on the body. Size, efficiency and power consumption were weighed heavily when selecting components. The DSP and amplifier are to be implemented on a PCB with their peripheral components. Without the peripheral components the combined area of the DSP and amplifier is below 1.5 cm^2 for either amplifier. The moving coil speaker is larger and also requires a cabinet to function properly, resulting in a size of about 5 x 5 x 5 cm. The piezo might be better suited to wear on the body with its 32.4 x 33.5 x 9.7 mm dimensions.
The system can handle inputs in the range of 50 Hz to 4 kHz without issue due to the sufficiently high sample rate. The speaker and amplifiers have been selected such that the sound intensity at 0.3 m is 85 dB or higher. However, since it was found that the most information in speech is found above 500 Hz, little priority was given to creating an output bandwidth that starts at 50 Hz. This was also due to the fact that speakers with

such range would be rather large. The chosen speakers have frequency responses that are well suited to accurately display the voice.
To limit power consumption and heat dissipation in the system, high efficiency and low power components were given priority. With the speaker that consumes 1 W nominally and 1.2 W maximum, the class D amplifiers that have an efficiency up to 90% and the low power DSP that consumes 51.7 mW during typical use, the power consumption stays well within acceptable range. The temperature of the device where it touches the skin depends on the casing that will be used. That topic is beyond the scope of this thesis, but choosing the efficient option will have impacted heat generation.

Signal processing requirements

The pitch shifting algorithm has been designed to meet the signal processing requirements. As mentioned above, an optimal shifting factor k has been determined using the STOI and SIIB algorithm. However, this factor will vary per patient. During this research there was only access to the audio files of one patient. The optimised factor for this patient might not work for the next. Moreover, it is difficult to decide if the intelligibility of the LP voice increased. The engineers who developed the pitch shifting algorithm are biased since they know the testing phrases by heart. Therefore it is difficult to decide if the intelligibility increased and if the shifted voice resembles a more healthy voice. And if so, the improvements are minimal. This could be due to the fact that the pitch is not the only aspect of the voice that determines intelligibility.
In other words, there is still a long way ahead. More techniques are required to improve the intelligibility and to create a more healthy voice. In Chapter 10 several future research topics will be discussed.

Costs requirements

The costs should not exceed the €100 as given in the statement of requirements. The selected components of the output part of the TESSA, described in this thesis, have a total costs of: €1,22 + €4,34 + €3,83 = €9,39. These are respectively the amplifier, speaker and the DSP. The total costs for the battery and microphone, which are not described in this thesis, are respectively: €28,00 + 2· €1,52 = €31,04. This brings the total costs of the entire system to: €9,39 + €31,04 = €40,43. This does not include the costs related to the assembly of and components required for the PCB implementation. Also, since no air pressure sensor was chosen, the cost of this feature is not included either. However it is not expected that these costs would bring the total above the €100 threshold.

Preferences

The proposed system meets most preferences. The latency stays below the 15 ms mark, at which the original and enhanced speech would be notably out of sync. The parts of the system have a combined cost lower than €100 when bought in bulk and the size has been kept as small as possible while still meeting the other requirements.
The naturalness of the voice is subjective as discussed above, but more work could be done to improve in this aspect. The volume regulation by air pressure has not been completely worked out but the main principles have been described. Further implementation is needed to meet this preference and give the user an intuitive way to control the volume of the device.

9 | Conclusion

The findings of the project can largely be divided into two distinct subjects. The selection and integration of the different hardware components that will amplify the voice, and the design of the algorithms that will shift the pitch in real-time. Judging by theory and simulations, the amplification and playback of the voice signal are sufficient while remaining in the boundaries set by the statement of requirements. While real-world testing remains a necessity to verify the functioning of the design and some additional features are not yet fully implemented, this part of the project can be viewed as a success.
The Ocean algorithm is used to shift the pitch of the voice and it does so successfully. However, the goal of the pitch shift is to make the voice more intelligible and sound more natural. Whether this effect has been achieved is debatable and any positive effect is unfortunately quite marginal. Shifting the pitch might be a part of the solution to let LP sound more natural, but more adaptions to voice should be considered.
The first steps in the design process have been taken. There is a hardware implementation that can be used as a basis for further research and the pitch shift could be part of the speech enhancement software although more work is needed in that respect. The following and final chapter will give recommendations for future steps in the design process.

10 | Recommendations

10.1. Implementation of unfinished features

Volume control

The volume control feature has unfortunately not been fully implemented. The proposed system uses air pressure in the throat of the user to control the volume which makes it intuitive and hands free. It is therefore recommended to use this method in further stages of development. The choice of the optimal sensor has not been part of this thesis and thus remains a subject for further research. Because it will be placed so close to or even inside of the body, safety is very important. Ideally the sensor would be passive and sensitive enough to detect small variations to correctly display inflections in the voice. The input voltage range of the analogue input pins of the DSP is between 0 and 3.6 V. The signal of the sensor might have to be amplified in order to be interpreted correctly by the DSP. It should also be noted that not every user has the same range of air pressure levels when speaking. There must be a way for the user to tweak the settings to achieve the best volume control for their voice.
The interface between the DSP and amplifier has been described in Chapter 6. It is known which pins should be addressed and which signals are required to change the amplification factor for each of the chosen amplifiers. However, as of yet there is no program that can be put on the DSP to process the pressure sensor input. Such a program could be relatively simple and use a switch statement in which the input of the sensor determines the case. Each case could then describe a different output towards the amplifier. It will require real world tests to tweak this program to match the air pressure with the right level of amplification.

Resonant frequency

Speakers have a frequency at which the component and its cabinet will resonate. This can result in unwanted ringing sounds and a peak in an otherwise flat frequency response [20]. For the chosen moving coil speaker this frequency is 550 Hz, the resonant frequency of the chosen piezo speaker is not mentioned in the data sheet. The unwanted behaviour can be prevented using filters. The resonant frequency that causes the behaviour can be eliminated using a notch filter that only stops a very narrow band of the signal [3]. The characteristics of such a filter depend on the electrical and mechanical driver of the speaker, Q_es and Q_ms respectively. These values are not mentioned in the data sheets, but can be determined by measuring the frequency response of the load impedance of the speaker. Unfortunately it was not possible to find these characteristics during the course of this thesis due to the unavailability of hardware and time constraints. A method of finding the driver factors is described in Appendix F.

10.2. Shifting alternatives

Roller's Algorithm

As mentioned in Chapter 3, the Roller's algorithm is another method of pitch shifting an audio signal. It could be used as an alternative for the Ocean algorithm described in this thesis, although hardware limitations need to be further investigated. The current calculations on DSP performances show some room for more heavily computing power performance and is therefore worth investigating.
As opposed to the Ocean algorithm, Roller's lives only in the time domain and actually uses many frequency shifting operations instead of one pitch shifting operation. The basis of the algorithm is a large bank of IIR filters to separate certain frequency bands. Then each frequency band is frequency shifted by a different amount, chosen such that the resulting spectrum is a scaled version of the original [17]. The performance of the algorithm is very good in terms of latency, achieving results around 5 ms. However, the authors state that the algorithm is implemented in Java and runs on a 2.4 GHz Pentium processor. Presumably this means an Intel Pentium 4 from 2002 is used. This single core processor runs at 2.4 GHz and has a 512 kB cache. In the paper there is no mention of other components such as a dedicated sound card or extra RAM. In a different paper where the same authors discuss the implementations of audio processing algorithms in Java, the Realtek SoundBlaster card is mentioned as the dedicated sound card [16].
At the time it took the processing power of a PC to run the algorithm. Current DSP technology has not quite reached the same clock speeds. However, the multi-core TI DSPs of the C6xxx series might be able to run the

algorithm as well as a Pentium 4 could. Although the highest clock speed of this line of DSPs is 1.4 GHz, they can have up to 12 cores. The processing capabilities do come at the cost of increased power consumption. For example, the 66AK2H14 can consume up to 12 W of power. Future researchers can investigate whether there are DSPs that can run the Roller's algorithm while not being too power hungry.

10.3. Speech synthesis

AR model

Speech can be described as an auto regressive model, meaning the next value can be predicted using previous values. When speech is produced, an excitation signal from the lungs is sent through the vocal chords and the mouth which alter the sound of the signal. These can be seen as filters applied to the signal, the response of which will be called $h_2(n)$ to conform with the used literature. This concept is used in telephony to model speech. The speech signal is sent through a filter that is the inverse of the filter formed in the throat and mouth. The result is a parametric description of speech that can be easily sent of telephone lines. At the receiver, the speech is synthesised using the incoming signal and a filter with response $h_2(n)$ [11].

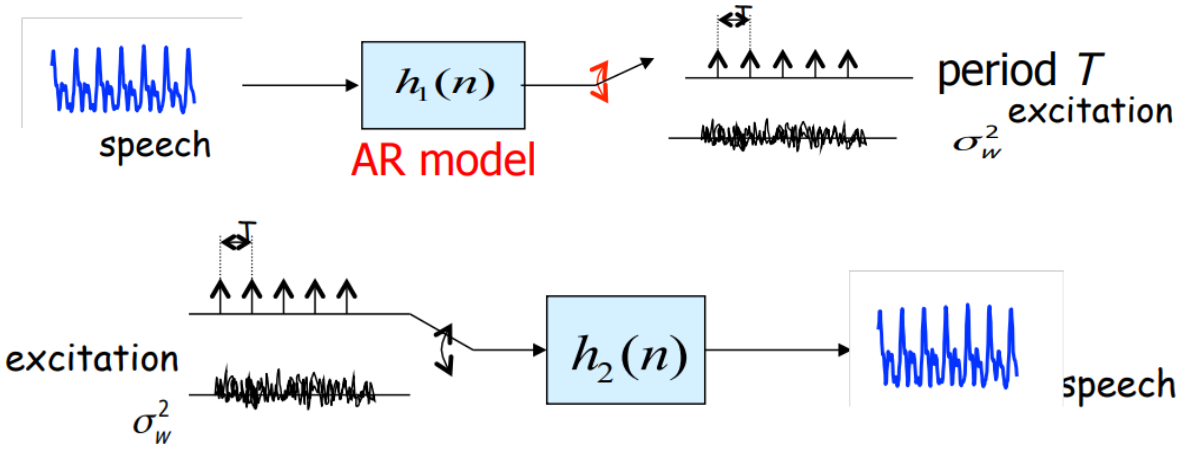


Figure 10.1: Use of AR model in telephony. (Above: sender, Below: receiver) [11]

Due to time constraints, this path remains unexplored in this thesis. Future research should be done into this form of speech synthesis and how it could be used to improve the user's voice. If the parameterisation of speech is successful it could be very simple to increase the fundamental frequency. However, it is known that characterising an LP voice can be more challenging than it is for healthy voices and not all current methods function well when applied to an LP [43]. This issue has also been encountered when trying to track the fundamental frequency for filtering purposes in the other subsystem.

Matsui

There has been some research into using speech synthesis as a solution for the problems LP face. A paper by Matsui et al. describes an approach that analyses the speech and synthesises an improved speech signal in real time. The process starts by dividing the speech signal using a band-pass filter and a high pass filter. The high frequencies are connected directly to the output, whereas the other frequencies are used for analysis.
The first point of analysis is to determine whether the input speech is voiced or unvoiced using a power threshold. Voiced signals are windowed and using linear predictive coding (LPC) and formant extraction, the peaks in the frequency domain are found. The pitch of the signal is also detected and using speech fragments of healthy voices, a new voice signal is synthesised. A block diagram of the whole process is shown in Appendix F.1.
The process can be implemented on a TI DSP, even in 2002 when this paper was published. The results are very positive as the test panel decided the synthesised speech was more intelligible and also some of the voice defects such as choppiness and strain were perceived as less prominent [21].
The most complete solution for improving LP voices might very well be based on this design. The elaborate

analysis-synthesis approach is beyond what could be completed in the 10 weeks of research for this thesis, but future stages of development are urged to look at this method of speech enhancement as it displays very promising results.

10.4. Further pitch shift testing

The pitch shifting algorithm is used to increase the intelligibility and the naturalness of the voice. These are used to adjust the pitch shift in this thesis. For further research, other methods can be utilised to measure these parameters.

Confidence in current method

The intelligibility of the audio fragments are measured with the STOI and SIIB algorithms until now. These are complex mathematical algorithms. The outcomes are difficult to relate to a certain confidence, since the outcomes do not have a clear meaning and unit. Although both testing methods provide same outcomes, it is not very sure whether the results of the STOI and SIIB algorithms correspond to the most intelligible result for the human ear. Since the device is made for humans, this is very important as well. Moreover, the testing audio fragments which are used are only based on one laryngectomised person, with spoken audio fragments from before the operation and after. The voice defects caused by laryngectomy are different per person. The fact that the pitch of the voice is lower and other defects are roughly the same for all patients, but to which extent these defects appear will differ. The audio fragments used in the audio fragments consist of just a few spoken sentences. The measurement results on the individual sentences provide almost the same outcome. However, there are not enough audio fragments available in order to create an extensive analysis on the entire speech spectrum to draw the conclusions for the ideal speech enhancement.

Pitch shift test recommendations

Further research can be done on intelligibility measurement techniques, although STOI and SIIB should be sufficient. Since the current testings are only based on one patient, it is recommended to record more audio samples from other patients in order to make a better analysis on the defects caused by laryngectomy. It is very important for further testing of the speech enhancement to record the new samples also with a wider variety of sentences. This means it should include all kinds of tones and letters which correspond with normal voice. The current audio fragments already consist of sentences commonly used in speech therapy. By recording a wider variety of this kind of sentences a good speech analysis could be created. These recorded need to be time aligned in order to make it suitable for direct testing with the STOI and SIIB algorithms. This can be achieved in several ways. For example, if the voice of a patient is recorded before the laryngectomy, the patient could speak the same text with an headphone in order to sync the new recording with the one before the operation. The measurements on intelligibility are now based on mathematics rather than human opinions. Whether the voice is more pleasant to listen to is determined by just a few people and may be biased as well since they are involved in the pitch shifting implementation. Therefore unbiased test panels could be used to test the intelligibility and naturalness of the result. A possible difficulty could be finding the optimal trade-off between intelligibility and naturalness, since these two characteristics do not always correlate with each other as noticed during the testing with the STOI algorithm. In order to test the filters and to make the pitch work in noisy environments, the recordings should be done in a noisy environment or with artificial additive noise. Since the voice defects are different for each patient, the final system should be designed in such a way that it could be personalised for each patient. The shifting factor must be adjusted for each user. Therefore it is recommended to make a standardised measurement setup which includes all recommendations described above.

10.5. Further hardware testing

Speaker selection

As described throughout this report, a moving coil and piezo speaker have been considered. Both devices have their pros and cons. The corona pandemic restricted the possibility of physically testing. Instead, both devices' characteristics have been simulated, as described in Chapter 7 but did not provide enough information to base the decision upon. Real world testing is required to determine if a piezo speaker can deliver the required sound pressure and quality. The size of the piezo makes it favourable over the moving coil, although the size of the moving coil speaker could turn out to be functional when assembling the voice system. When

testing the speakers in the real world, the aim for the speakers is to produce a sound pressure level (SPL) of 85 dB at 0.3 m, especially in the range of 500 Hz-4 kHz. A typical test setup could be to record multiple audio fragments at this distance from each speaker using a high quality audio microphone and amplifier. The recorded signal can be analysed using MatLab. The STOI and SIIB algorithm, described in Section 7.1 and 7.1 respectively, can be used to test the intelligibility of the recorded sound by selecting the original fragment as 'clean speech' and the recorded signal as 'degraded speech'. The results of these algorithms can be compared to determine which speaker has a higher intelligibility rating.

DSP

First of all, the described functionalities and algorithms should be converted into C/C++ programming language to be able to run on the DSP. This can be done via CodeComposerStudio, a software development tool designed by TI for their DSPs. After successfully implementing the code, it can be loaded on the DSP. An evaluation module kit can be used to physically test the code on the DSP using CodeComposerStudio again.
The estimation of the power consumption is only based on the documentation that TI has provided. This includes a model of typical use for different clock speeds and the power used by the HWA FFT module when computing different lengths of FFT. However, the typical power consumption while running the designed algorithms and applying the designed filters will likely differ from the typical consumption modelled by TI. Furthermore, the power consumed by the FFT operations might not be an accurate indicator of how much power the different algorithms consume. There are many other operations that are executed for each frame of which the required energy is unknown. It is therefore recommended to measure the power consumption of the DSP while it is running to get a more accurate view. The findings could lead to a change in the power budget and allow for a smaller or cheaper battery to be placed in the system.

Amplifier

The speaker selection greatly influences the amplifier selection. After determining which speaker shall be implemented, its accompanying amplifier can be tested. Several tests can be run on the amplifier, such as a SPL test and intelligibility test at 0.3 m. As described above, the STOI and SIIB algorithm can be used to determine in which extent the amplifier deteriorates the intelligibility.
Finally, the efficiency of the amplifier can be tested. An appropriate test signal is suggested in [15], instead of sine waves. The crest factor, i.e. the ratio between the peak- and RMS value of a signal in decibels, of sine waves equals 3 dB while the crest factor for audio tracks has a mean value of 15 dB. The suggested test signal is more representative of audio tracks than simple sinusoids yet has a fundamental frequency which makes it easier to tell when the signal starts to distort by looking at the waveform with an oscilloscope.

Bibliography

[1] From speech physiology to acoustics 1:air pressure and sources of sound. URL http://www.phon.ox.ac.uk/jcoleman/sources.htm#:~:text=Pressure%20and%20air%20flow%20during,1200%20Pa%20during%20speech%20production.

[2] A SHORT-TIME OBJECTIVE INTELLIGIBILITY MEASURE FOR TIME-FREQUENCY WEIGHTED NOISY SPEECH Signal Information & Processing Lab , 2628 CD Delft , The Netherlands Oticon A / S 2765 Smørum , Denmark. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 4214–4217, 2010. ISSN 1175-5326. URL http://cas.et.tudelft.nl/pubs/Taal2010.pdf.

[3] All about circuits. Resonant filters. URL https://www.allaboutcircuits.com/textbook/alternating-current/chpt-8/resonant-filters/.

[4] *Digital Input, Mono 2 W, Class-D Audio Power Amplifier*. Analog Devices, 2012. URL https://www.analog.com/media/en/technical-documentation/data-sheets/SSM2529.pdf. Rev. 0.

[5] James P. Anthony Charles G. Reed Mark I. Singer Daniel G. Deschler, E. Thomas Doherty. Tracheoesophageal voice following tubed free radial forearm flap reconstruction of the neopharynx. *Annals of Otology, Rhinology Laryngology*, 103:929–936, 1994.

[6] MP Divakar. Surface temperatures of electronics products: Appliances vs. wearables, 2016. URL https://www.electronics-cooling.com/2016/09/surface-temperatures-of-electronics-products-appliances-vs-wearables/.

[7] PH.D. P. DELAERE M.D. PH.D. J. WOUTERS PH.D. P. UWENTS M.D. F. DEBRUYNE, M.D. Acoustic analysis of tracheo-oesophageal versus oesophageal speech. *The Journal of Laryngology and Otology*, 108: 325–328, 1994.

[8] *Integrated Silicon Pressure Sensor On-Chip Signal Conditioned,Temperature Compensated and Calibrated*. Freescale Semiconductor, 5 2010. URL https://docs.rs-online.com/63e4/0900766b814cadd1.pdf.

[9] Victoria State Government. Noise, 2018. URL https://www.education.vic.gov.au/school/students/beyond/Pages/noise.aspx#:~:text=Normal%20conversation%20is%20about%2060,shout%20to%20make%20yourself%20heard.

[10] Sourav Gupta. Classes of power amplifiers, 2018. URL https://circuitdigest.com/tutorial/classes-of-power-amplifier-explained.

[11] Richard C. Hendriks. Stochastic processes- lecture 7: signal processing supplement - filtering stochastic processes. University Lecture, 2020.

[12] N. Juillerat; B. Hirsbrunner. Low latency audio pitch shifting in the frequency domain. *2010 International Conference on Audio, Languange and Image Processing*, pages 16–24, 2010.

[13] Mark Hughes. How piezoelectric speakers work, 2016. URL https://www.allaboutcircuits.com/technical-articles/how-piezoelectric-speakers-work/.

[14] Measurement Specialities Inc. Efficiency comparison between moving-coil and piezo film speakers, 2000. URL https://www.metrolog.net/files/piezofilm_speaker_en_metrolog.pdf.

[15] Niels Elkjær Iversen, Arnold Knott, and Michael A. E. Andersen. Efficiency of switch-mode power audio amplifiers - test signals and measurement techniques. In *Proceedings of the 140th Audio Engineering Convention Convention*. Audio Engineering Society, 2016. 140th International Audio Engineering Society Convention ; Conference date: 04-06-2016 Through 07-06-2016.

[16] N. Juillerat and S. Schubiger-Banz. Real-time, low latency audio processing in java. pages 99–102, 2007. ICMC 2007 - Proceedings of the International Computer Music Conference,.

[17] S.; Arisona S.M. Juillerat, N.; Schubiger-Banz. Low latency audio pitch shifting in the time domain. pages 29–35, 2008. ICALIP 2008 - Proc. of the IEEE International Conference on Audio Language and Image Processing, Shanghai, China; pp. 29-35.

[18] A. Keltz. Latency and digital audio systems. URL http://whirlwindusa.com/support/tech-articles/opening-pandoras-box/.

[19] Martin Logan. Electrostatic (esl) theory. URL https://www.martinlogan.com/en/electrostatic-esl-theory.

[20] Mareo Lopez. The importance of a speaker's resonant frequency, 2015. URL https://www.proaudioland.com/news/the-importance-of-a-speakers-resonant-frequency/.

[21] Kenji Matsui, Noriyo Hara, Noriko Kobayashi, and Hajime Hirose. Enhancement of esophageal speech using formant synthesis. Technical report.

[22] Paul McGowan. How a ribbon speaker works, 2018. URL https://www.psaudio.com/pauls-posts/how-a-ribbon-speaker-works/.

[23] Trialsight Medical Media. Anatomy of the larynx, 2008. URL http://www.hawaiivoiceandspeechstudio.com/blog/archives/05-2016.

[24] Atos Medical. Esophageal speech. URL https://www.atosmedical.us/support/esophageal-speech/.

[25] Electronic Notes. Moving coil loudspeaker. URL https://www.electronics-notes.com/articles/audio-video/loudspeaker/moving-coil-speaker.php.

[26] Jan Onderka. Pitch shifting of audio signals in real time using stft on a digital signal processor. Bachelor's thesis, Faculty of Information Technology CTU in Prague, 2018.

[27] Christopher G. Tang Rachel Kaye and Chatherine F. Sinclair. The electrolarynx: voice restoration after total laryngectomy. *Med Devices*, 10:133 —- 140, 2017. doi: 10.2147/MDER.S133225. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5484568/.

[28] A. J. Roscoe and G. M. Burt. Comparisons of the execution times and memory requirements for high-speed discrete fourier transforms and fast fourier transforms, for the measurement of ac power harmonics. pages 40–45, 6 2011. 2nd IMEKO TC 11 International Symposium Metrological Infrastructure ; Conference date: 15-06-2011 Through 17-06-2011.

[29] SoundBridge. Piezoelectric speaker, 2019. URL https://soundbridge.io/piezoelectric-speaker/.

[30] Morten Stove. Facts about speech intelligibility, 2016. URL https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility.

[31] Johan Sundberg, Ronald Scherer, Markus Hess, Frank Müller, and Svante Granqvist. Subglottal pressure oscillations accompanying phonation. *Journal of Voice*, 27:411–21, 07 2013. doi: 10.1016/j.jvoice.2013.03.006.

[32] Kerry Hanssen Paul G. Beaudin Tanya L. Eadie, Philip C. Doyle. Influence of speaker gender on listener judgments of tracheoesophageal speech. *The Journal of Voice*, 22:43–57, 2008.

[33] *19-V_{PP} Mono Class-D Audio Amplifier for Piezo/Ceramic Speakers*. Texas Instruments, 12 2008. URL https://www.ti.com/lit/ds/symlink/tpa2100p1.pdf.

[34] *FFT Implementation on the TMS320VC5505, TMS 320 C5505, and TMS320C5515 DSPs*. Texas Instruments, 1 2013. Rev. B.

[35] *TMS320C5515 Fixed-Point Digital Signal Processor*. Texas Instruments, 10 2013. Rev. F.

[36] *TMS320C5515/14/05/04 SP Inter-IC Sound (I2S) Bus User's Guide*. Texas Instruments, 5 2014. Rev. B.

[37] *Power Estimation and Power Consumption Summary for TMS320C5504/05/14/15/32/33/34/35/45 Devices*. Texas Instruments, 4 2016. Rev. A.

[38] *TMS320C5505/15/35/45 schematic checklist*. Texas Instruments, 2 2019.

[39] Hunter E. J. Svec J. G. Titze, I. R. Voicing and silence periods in daily and weekly vocalizations of teachers. *The Journal of the Acoustical Society of America*, 121:469 —- 478, 2007. doi: https://doi.org/10.1121/1.2390676. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371399/.

[40] Ravit Cohen Mimran Tova Most, Yishai Tobin. Acoustic and perceptual characteristics of esophageal and tracheoesophageal speech production. 33:165–181, 03 2000.

[41] C.J. van As. Tracheoesophageal speech. a multidimensional assessment of voice quality. Technical report, 2001. URL https://pure.uva.nl/ws/files/3099184/143426_A01_201_006.pdf.

[42] Steven Van Kuyk, W Bastiaan Kleijn, and Richard C Hendriks. An Instrumental Intelligibility Metric Based on Information Theory. *IEEE SIGNAL PROCESSING LETTERS*, 25(1):115, 2018. doi: 10.1109/LSP.2017.2774250. URL http://ieeexplore.ieee.org.

[43] Begoña García Zapirain, Javier Vicente, B García, and J Vicente. Software for Measuring and Improving Esophageal Voices. Technical report, 2004. URL https://www.researchgate.net/publication/228441108.

A | Morphological maps

Subjects	Options			
Type	Dynamic (moving coil)	Dynamic (ribbon)	Electrostatic	
Shape	Low profile	Cone	Shoulder pad	
Number of speakers	1 (Full range)	2 (Mid + high)	3 (Low mid high)	
Location	On chest under clothes	On chest above clothes	On shoulder	

Figure A.1: Morphological map speaker

Subjects	Options			
Class	A	B	A/B	D
Signal type	Analog	Digital		
Volume control component	Potentiometer	Variable gain amplifier	Voltage controlled resistor	
Output channel	Mono	Stereo		
Handsfree Volume control	Air pressure	Voice volume		

Figure A.2: Morphological map amplifier

B | Component selection rating

B.1. DSP selection rating

DSP	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	Criterion 6	Criterion 7	Criterion 8	End score
C5504	1	1	1	-1	1	-1	1	1	11
C5515	1	1	1	1	1	0	1	1	20
BF527C	0	1	1	1	1	1	0	0	17
BF707	-1	1	1	1	1	0	0	1	15
Sharc21566	-1	1	1	1	-1	-1	0	0	5

Positief		
Negatief		
Neutraal		
	Weight factor	
Criterion 1	2	Input voltage +/- 0.3 V
Criterion 2	4	> 10 kB ROM
Criterion 3	4	> 100kB RAM
Criterion 4	4	Ability to perform FFT
Criterion 5	4	Power consumption
Criterion 6	1	Ease of implementation
Criterion 7	1	Ease of programming
Criterion 8	1	Price

Figure B.1: Score system of the available DSPs

B.2. Speaker selection rating

Speaker	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	Criterion 6	Criterion 7	Criterion 8	End score
midrange PRO	-1	1	1	1	-1	1	0	0	8
k57	1	-1	0	1	1	1	0	0	6
K16	-1	-1	1	1	1	-1	0	1	-3
PS95-8	1	0	-1	-1	1	1	0	-1	3
RS Pro	0	1	0	1	0	1	0	0	10
FRWS5R	1	1	0	-1	1	1	0	-1	9
FR7_4	1	0	-1	-1	1	1	1	0	5
K 50 WP	1	1	0	0	-1	1	0	0	8
SCS-32	-1	1	1	1	0	1	0	0	10
SPS-4640	0	1	1	0	0	1	0	0	10

Positive		
Negative		
Neutral		
	Weight factor	
Criterion 1	2	Frequency band
Criterion 2	4	Frequency response 2kHz - 4kHz
Criterion 3	2	Size
Criterion 4	2	Rated power
Criterion 5	2	Maximum power
Criterion 6	4	Minimum sound pressure level 1W @0.3m
Criterion 7	1	Input impedance
Criterion 8	1	Cost

Figure B.2: Score system of the available speakers

B.3. Amplifier selection rating

Amplifier	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	Criterion 6	Criterion 7	Criterion 8	End score
SSM4567	1	1	1	1	1	0	0	0	13
SSM2529	1	1	1	1	1	1	0	0	15
SSM2305	1	1	0	-1	1	1	0	0	11
PAM8405	1	1	0	-1	1	0	1	1	11
PAM8302A	1	1	0	-1	1	0	1	1	11
MAX98355A/B	1	1	1	-1	1	0	0	0	11
LM4675	1	1	0	1	1	1	0	1	14

Positief	
Negatief	
Neutraal	

	Weight factor	
Criterion 1	3	Frequency range
Criterion 2	4	Power to 8 Ohm load
Criterion 3	2	THD (1W to 8 Ohm)
Criterion 4	1	Built in volume control
Criterion 5	3	Efficiency
Criterion 6	2	SNR
Criterion 7	1	Gain
Criterion 8	1	Price

Figure B.3: Score system of the available amplifiers

B.4. Amplifier selection rating (Piezo)

Amplifier	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	Criterion 6	Criterion 7	Criterion 8	End score
LT3469-01	1	1	0	1	-1	0	0	0	7
MAX-9788-01	1	1	0	0	1	0	0	0	9
LM4960SQ-02	1	1	1	0	0	0	0	0	9
StepUpBTL-01	1	1	-1	1	0	0	0	-1	6
TPA2100P1	1	1	1	0	1	0	0	1	13

Positief	
Negatief	
Neutraal	

	Weight factor	
Criterion 1	2	Frequency range
Criterion 2	4	Ability to drive 66nF load
Criterion 3	3	Minimum supply voltage
Criterion 4	4	Output Vpp
Criterion 5	3	Efficiency
Criterion 6	1	Built in volume control
Criterion 7	1	Gain
Criterion 8	1	Price

Figure B.4: Score system of the available piezo amplifiers

C | MatLab scripts

C.1. Shifting function script

```
%Author: Tim Alers and Joris van Breukelen
clear all;
close all;

%% Reading the data
[y,Fs] = audioread('post 1 jaar geknipt.mp3'); %reading out data
y = resample(y,16000,44100); %Resampling the data
Fs = 16000;

%% Set up data
samplesize = length(y);
f_axis = [0:1/(samplesize-1):1]*Fs;

k = 1.1; %Factor of frequency shifting
m = 1;
osamp = 1; %oversampling rate
framesize_duration = 1; %In seconds
framesize = floor(Fs*framesize_duration);
frame_number = 1;

%% Ocean Algorithm
Y = fft(y(:,1));

%multiplier = 2*Pi*p/OmN
multiplier = (2*pi*frame_number)/(osamp*m*framesize);

%b = mka + 0.5
for bin = 1:round((samplesize))
    new_bin = floor(m*k*bin+0.5);
    if new_bin >= 0 && new_bin < samplesize/2
        oscVar = (new_bin - (m * bin)) * multiplier;
        shifted_Y(new_bin) = Y(bin)*exp(j*oscVar);
        shifted_Y(samplesize-new_bin+1) = Y(bin)*exp(j*oscVar);
    end
end

%% Plot data
shifted_y = real(ifft(shifted_Y));

hold on;
plot(f_axis/1000,abs(shifted_Y));
plot(f_axis/1000,abs(Y));
legend( 'Shifted frequency spectrum','Original frequency spectrum');
xlabel('Frequency [kHz]')
ylabel('Amplitude')
title('Frequency spectra of shifted frequencies by a factor of 1.1')
xlim([0 8]);

%% Play Audio
```

```
player = audioplayer(shifted_y, Fs);
% play(player);

%% Save audiofile
time_seconds = 5;
% audiowrite('audio_before_shifted.wav',y(1:time_seconds*Fs,:),Fs)
```

C.2. Shifting function of entire project

```
function [shifted_Y, frame_number] = frequency_shifting(Y,m,k,osamp,frame_number)
%% Ocean Algorithm
framesize = length(Y); %Defining framesize
multiplier = (2*pi*frame_number)/(osamp*m*framesize); %multiplier = 2*Pi*p/OmN

for bin = 1:round((framesize)/2)
    new_bin = floor(m*k*bin+0.5); %Assigning new frequency bin (b = mka + 0.5)
    if (new_bin >= 0 && new_bin < framesize/2) % Only half of frequency spectrum
        because of symmetry
        oscVar = (new_bin - (m * bin)) * multiplier; %Phase correction
        shifted_Y(new_bin) = Y(bin)*exp(j*oscVar); %Freq shift en phase correction
            on first half
        shifted_Y(framesize-new_bin+1) = Y(bin+1)*exp(j*oscVar); %Freq shift en
            phase correction on second half
    end
end
frame_number = frame_number + 1; %Updating frame number
shifted_Y = shifted_Y'; %Transposing output for filter group
end
```

C.3. Intelligibility tests

```
clear all;
close all;

%% Reading in audio files

[pre,Fs1] = audioread('pre.mp3'); %Audio fragment before operation
% pre = resample(pre,16000,44100);
[post,Fs2] = audioread('post.mp3');%Audio fragment after operation
% pre = resample(pre,16000,44100);

pre = pre(:,1); %Selecting right channel
post = post(1:length(pre),1); %Selecting right channel and making it the right size

%% Testing all possible shifting factors
n = 1;
Fs = 44100;
for shifting_factor = 1:0.01:1.3
    y = post; %Recovering original signal
    read_buffered_audio_window_test; %Applying shift operation
    b(n,2) = stoi(pre,inverse_shifted,Fs); %Applying SIIB algorithm
    b(n,1) = shifting_factor; %Storing corresponding shifting factor
    n = n + 1;
end

```

```
25 %% Plotting results
26
27 plot(b(:,1),b(:,2))
28
29 title('STOI outcome for different shifting factors')
30 xlabel('Shifting factor')
31 ylabel('Intelligibility')
```

D | DSP documentation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
P	EM_DQM1	DV_{DDEMIF}	DV_{DDIO}	LCD_CS0_E0/SPI_CS0	LCD_RW_WRB/SPI_CS2	LCD_D[0]/SPI_RX	LCD_D[2]/GP[12]	DV_{DDIO}	LCD_D[5]/GP[15]	LCD_D[7]/GP[17]	LCD_D[9]/I2S2_FS/GP[19]/SPI_CS0	LCD_D[11]/I2S2_DX/GP[27]/SPI_TX	LCD_D[13]/UART_CTS/GP[29]/I2S3_FS	LCD_D[15]/UART_TXD/GP[31]/I2S3_DX
N	EM_A[15]/GP[21]	EM_SDCKE	LCD_EN_RDB/SPI_CLK	LCD_CS1_EN1/SPI_CS1	LCD_RS/SPI_CS3	LCD_D[1]/SPI_TX	LCD_D[3]/GP[13]	LCD_D[4]/GP[14]	LCD_D[6]/GP[16]	LCD_D[8]/I2S2_CLK/GP[18]/SPI_CLK	LCD_D[10]/I2S2_RX/GP[20]/SPI_RX	LCD_D[12]/UART_RTS/GP[28]/I2S3_CLK	LCD_D[14]/UART_RXD/GP[30]/I2S3_RX	DV_{DDIO}
M	EM_A[14]	EM_D[5]	EM_SDCLK	$\overline{EM\_CS3}$	EMU1	TCK	TDO	XF	$\overline{TRST}$	MMC0_D1/I2S0_RX/GP[3]	MMC0_CMD/I2S0_FS/GP[1]	MMC1_D1/I2S1_RX/GP[9]	MMC1_CLK/I2S1_CLK/GP[6]	MMC1_D0/I2S1_DX/GP[8]
L	EM_A[13]	EM_A[10]	EM_D[12]	EM_D[4]	CV_{DD}	EMU0	TDI	TMS	MMC0_D0/I2S0_DX/GP[2]	MMC0_CLK/I2S0_CLK/GP[0]	MMC0_D3/GP[5]	MMC0_D2/GP[4]	MMC1_D3/GP[11]	MMC1_CMD/I2S1_FS/GP[7]
K	EM_A[12]/(CLE)	EM_A[11]/(ALE)	EM_D[14]	EM_D[13]	EM_D[6]	EM_WAIT3	DV_{DDIO}	V_{SS}	V_{SS}	CV_{DD}	V_{SS}	DV_{DDIO}	V_{SS}	MMC1_D2/GP[10]
J	EM_A[8]	EM_A[9]	EM_A[20]/GP[26]	EM_D[15]	DV_{DDEMIF}	CV_{DD}	V_{SS}	V_{SS}	V_{SS}	RSV1	RSV2	USB_VBUS	USB_V_{DD1P3}	USB_DM
H	$\overline{EM\_WE}$	EM_A[7]	EM_D[7]	EM_WAIT5	DV_{DDEMIF}	V_{SS}	DV_{DDEMIF}	CV_{DD}	USB_V_{SSA1P3}	USB_V_{DDA1P3}	USB_V_{SSA3P3}	USB_V_{DDA3P3}	USB_V_{SS1P3}	USB_DP
G	EM_WAIT4	EM_A[18]/GP[24]	EM_D[0]	EM_A[19]/GP[25]	DV_{DDEMIF}	V_{SS}	V_{SS}	USB_V_{DDPLL}	USB_R1	USB_V_{SSREF}	USB_V_{SSPLL}	USB_V_{DDOSC}	USB_MXI	USB_MXO
F	EM_A[6]	EM_A[17]/GP[23]	EM_D[2]	EM_D[9]	DV_{DDEMIF}	CV_{DD}	DV_{DDIO}	DV_{DDRTC}	V_{SS}	V_{SS}	USB_V_{SSOSC}	USB_LDOO	LDOI	LDOI
E	EM_A[2]	EM_A[16]/GP[22]	EM_D[8]	$\overline{EM\_OE}$	EM_D[1]	DV_{DDEMIF}	$\overline{INT1}$	WAKEUP	V_{SS}	DSP_LDOO	V_{SS}	V_{SS}	V_{SS}	V_{SS}
D	EM_A[5]	EM_A[3]	EM_D[10]	EM_D[3]	EM_WAIT2	$\overline{RESET}$	V_{SS}	RTC_CLKOUT	V_{SSA_PLL}	GPAIN0	V_{SS}	$\overline{DSP\_LDO\_EN}$	RSV16	RSV3
C	EM_A[4]	EM_A[1]	$\overline{EM\_CS4}$	EM_D[11]	$\overline{EM\_CS2}$	$\overline{INT0}$	CLK_SEL	CV_{DDRTC}	V_{SSRTC}	V_{DDA_PLL}	GPAIN3	RSV0	RSV5	RSV4
B	EM_BA[1]	EM_A[0]	$\overline{EM\_CS0}$	$\overline{EM\_SDCAS}$	EM_DQM0	EM_R/$\overline{W}$	SCL	SDA	RTC_XI	V_{SSA_ANA}	GPAIN2	LDOI	BG_CAP	V_{SSA_ANA}
A	EM_BA[0]	DV_{DDEMIF}	$\overline{EM\_CS5}$	$\overline{EM\_CST}$	DV_{DDEMIF}	$\overline{EM\_SDRAS}$	CLKOUT	CLKIN	RTC_XO	V_{DDA_ANA}	GPAIN1	ANA_LDOO	V_{SS}	V_{SS}

Figure 2-2. Pin Map

Figure D.1: DSP pin map as taken from the TI manual [35]

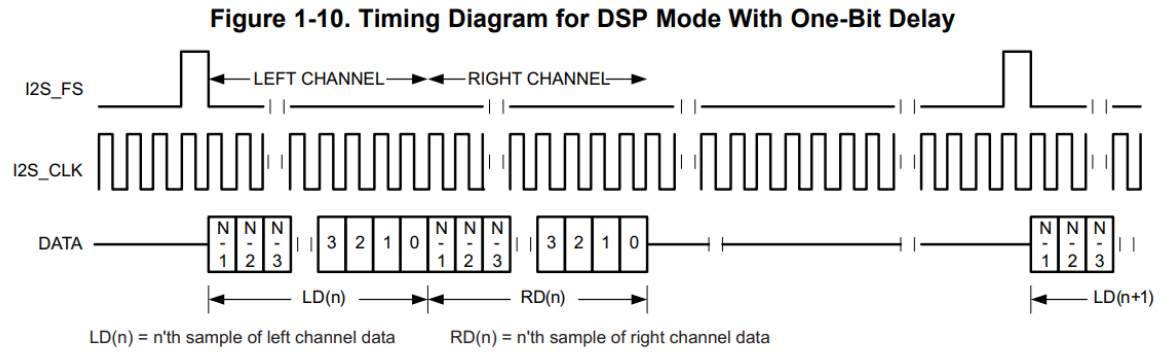


Figure D.2: Timing diagram for DSP mode with one-bit delay as taken from the TI manual [36]

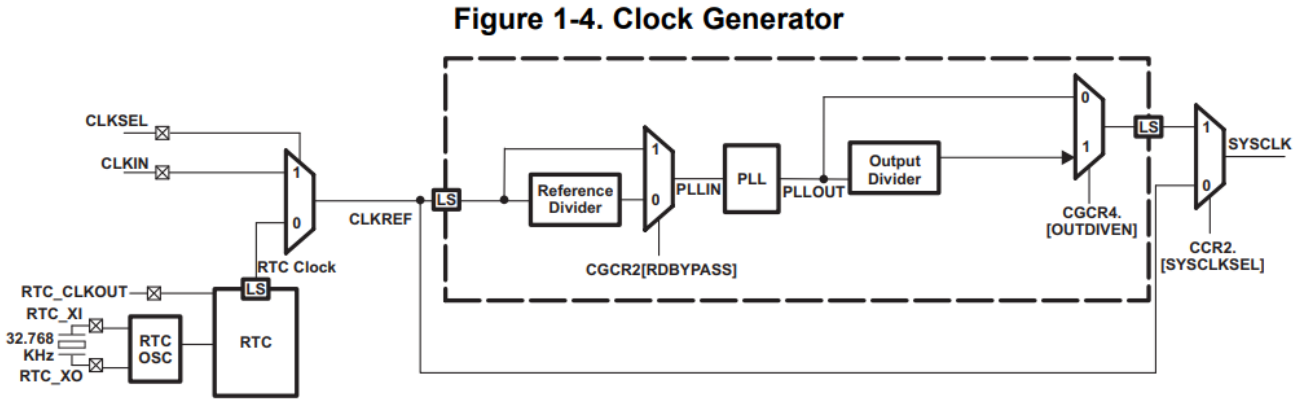


Figure D.3: Schematic of the system clock generator as taken from the TI manual [36]

Table 3. FFT Performance on HWAFFT vs CPU (Vcore = 1.05 V, PLL = 60 MHz)

Complex FFT	FFT with HWA		CPU (Scale)		HWA versus CPU	
	FFT + BR [1] Cycles	Energy/FFT (nJ/FFT)	FFT + BR [1] Cycles	Energy/FFT (nJ/FFT)	x Times Faster (Scale)	x Times Energy Efficient (Scale)
8 pt	92 + 38 = 130	23.6	196 + 95 = 291	95.1	2.2	4
16 pt	115 + 55 = 170	32.1	344 + 117 = 461	157.1	2.7	4.9
32 pt	234 + 87 = 321	69.5	609 + 139 = 748	269.9	2.3	3.9
64 pt	285 + 151 = 436	98.5	1194 + 211 = 1405	531.7	3.2	5.4
128 pt	633 + 279 = 912	219.2	2499 + 299 = 2798	1090.4	3.1	5
256 pt	1133 + 535 = 1668	407.2	5404 + 543 = 5947	2354.2	3.6	5.8
512 pt	2693 + 1047 = 3740	939.7	11829 + 907 = 12736	5097.5	3.4	5.4
1024 pt	5244 + 2071 = 7315	1836.2	25934 + 1783 = 27717	11097.9	3.8	6

[1] BR = Bit Reverse

Table 4. FFT Performance on HWAFFT vs CPU (Vcore = 1.3 V, PLL = 100 MHz)

Complex FFT	FFT with HWA		CPU (Scale)		HWA versus. CPU	
	FFT + BR [1] Cycles	Energy/FFT (nJ/FFT)	FFT + BR [1] Cycles	Energy/FFT (nJ/FFT)	x Times Faster (Scale)	x Times Energy Efficient (Scale)
8 pt	92 + 38 = 130	36.3	196 + 95 = 291	145.9	2.2	4
16 pt	115 + 55 = 170	49.3	344 + 117 = 461	241	2.7	4.9
32 pt	234 + 87 = 321	106.9	609 + 139 = 748	414	2.3	3.9
64 pt	285 + 151 = 436	151.3	1194 + 211 = 1405	815.7	3.2	5.4
128 pt	633 + 279 = 912	336.8	2499 + 299 = 2798	1672.9	3.1	5
256 pt	1133 + 535 = 1668	625.6	5404 + 543 = 5947	3612.9	3.6	5.8
512 pt	2693 + 1047 = 3740	1442.8	11829 + 907 = 12736	7823.8	3.4	5.4
1024 pt	5244 + 2071 = 7315	2820.6	25934 + 1783 = 27717	17032.4	3.8	6

[1] BR = Bit Reverse

Figure D.4: Tables on FFT performance as taken from TI manual [34]

E | Simulation amplifier

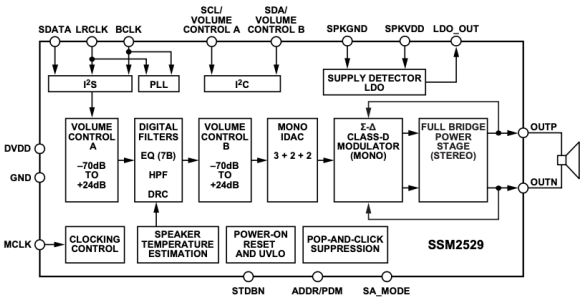


Figure E.1: SSM2529 digital amplifier functional block diagram

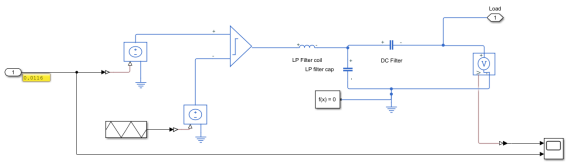


Figure E.2: Simulation scheme in Simulink of the class D amplifier

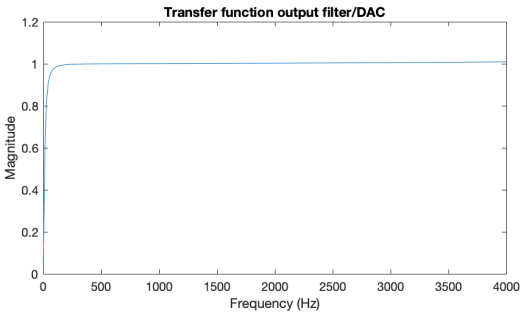


Figure E.3: Moving coil output filter response up to 4kHz

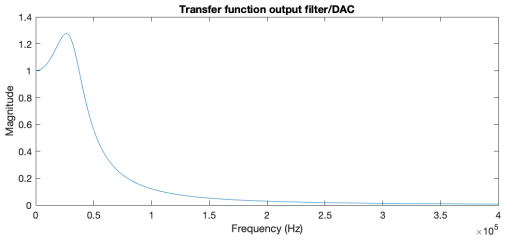
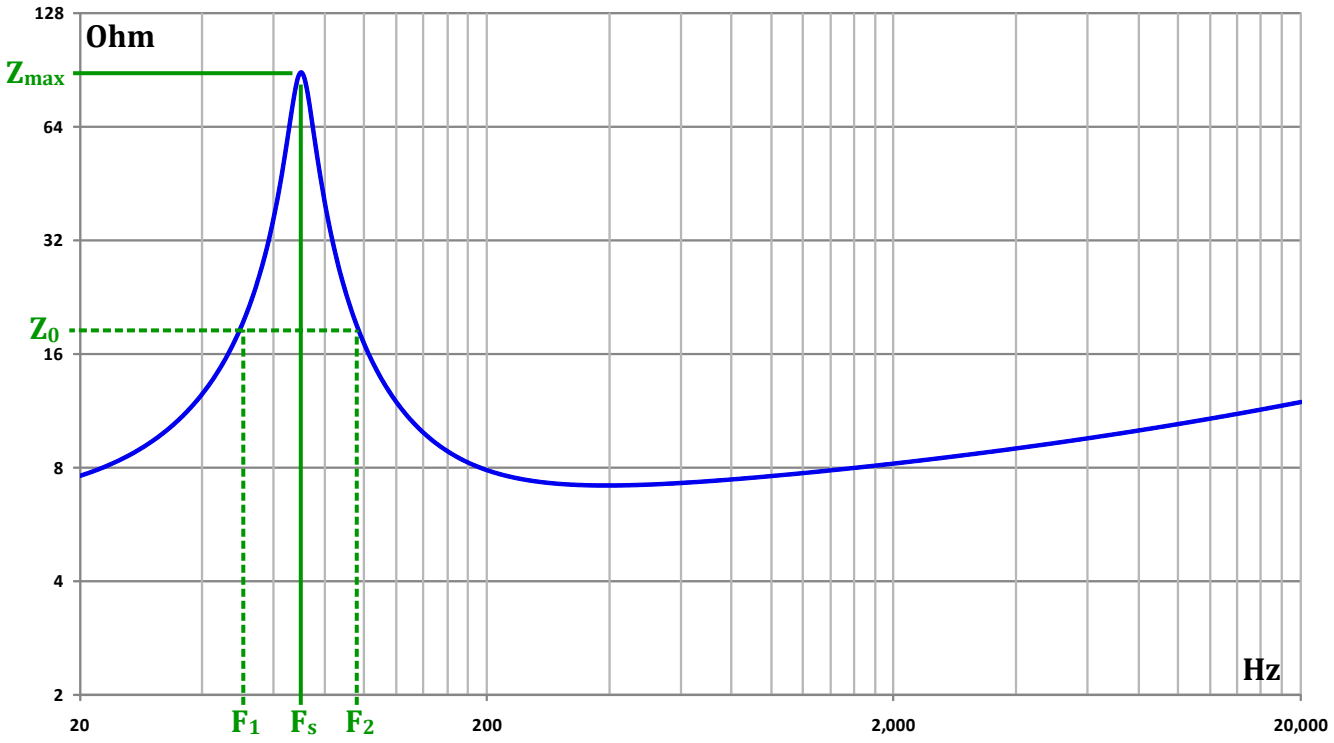


Figure E.4: Moving coil output filter response up to 400kHz

F | Measuring Q-values

Measuring/calculating driver Q values using the impedance curve

R_e	DC resistance of driver
F_s	Resonance frequency of driver
Z_{max}	Impedance at F_s
Z_0	Calculated impedance level for reading F_1 and F_2
F_1	Lower frequency where $Z=Z_0$
F_2	Upper frequency where $Z=Z_0$
Q_{ms}	Mechanical Q of driver
Q_{es}	Electricaal Q of driver
Q_{ts}	Total Q of driver



Qms

Step 1. Measure R_e of the driver

Step 2. Read the values F_s and Z_{max} from the impedance curve

Step 3. Calculate Z_0 as $Z_0 = \sqrt{R_e \times Z_{max}}$

Step 4. Read the values F_1 and F_2 from the impedance curve

Step 5. Calculate Q_{ms} as $Q_{ms} = \frac{F_s}{(F2-F1)} \times \sqrt{\frac{Z_{max}}{R_e}}$

Qes

Calculate Q_{es} as $Q_{es} = \frac{Q_{ms}}{\frac{Z_{max}}{R_e} - 1}$

Qts

Calculate Q_{ts} as $Q_{ts} = \frac{Q_{ms} \times Q_{es}}{Q_{ms} + Q_{es}}$

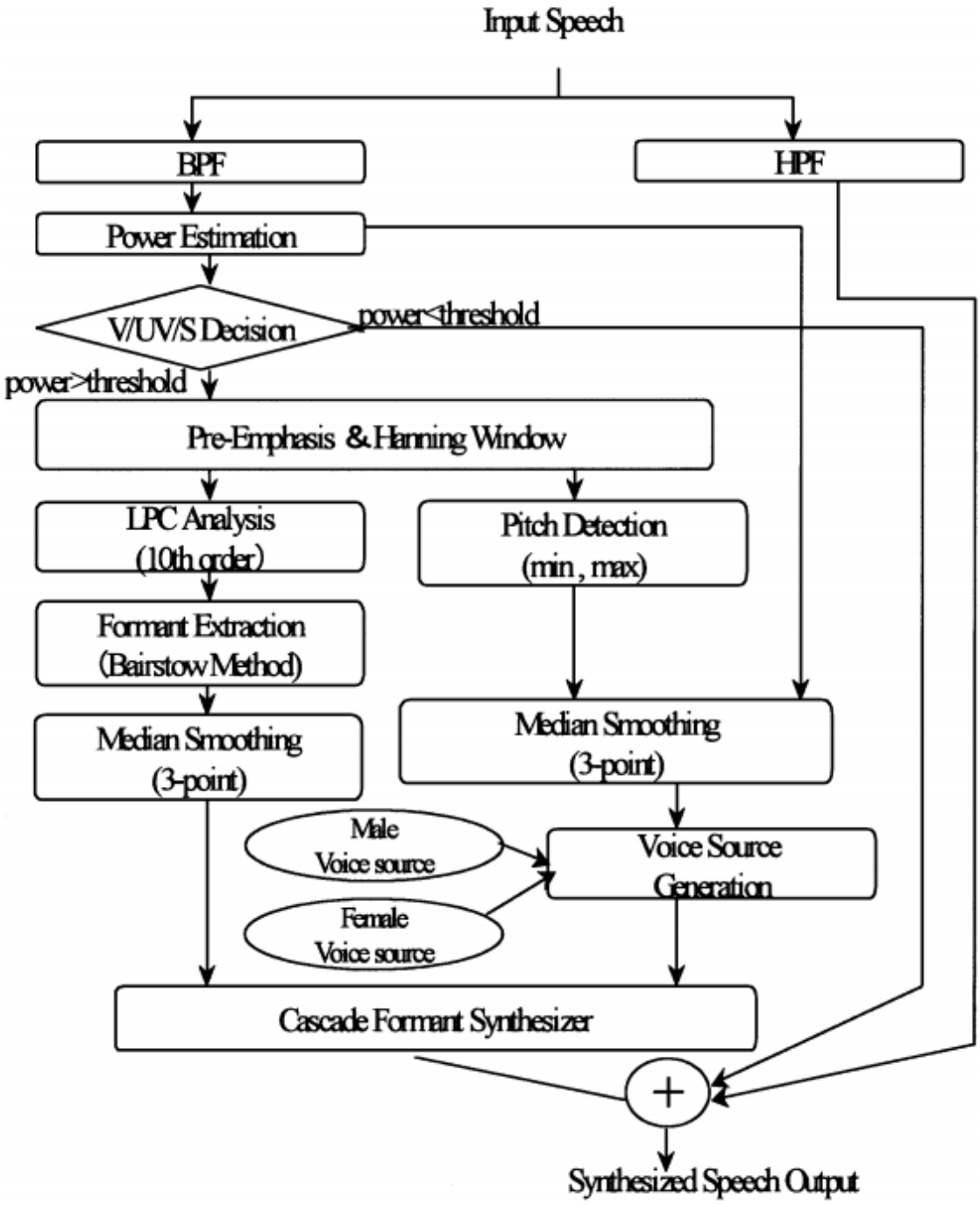


Figure F.1: Block diagram of the speech analysis-synthesis process [21]

G | Statement of requirements

Below a translation of the statement of requirements agreed upon with the client.

G.1. Assignment description

The voice amplification device is part of a second generation aid device for Laryngectomised People (LP) called the Exobreather. This device conditions air (filter, heat and moisturises) what would normally have been done by the nose mouth and throat cavities. Furthermore a hands free speech utility and a stoma fluid absorption utility are in the pipeline to reduce stoma maintenance to once a day. This device is worn beneath the clothing, below the stoma on the upper chest. The voice amplification device is, including all its peripheral equipment, part of the Exobreather. For this reason the complete design should be as compact as possible.
The system should be an all in one package with preferentially no loose parts. For that reason a microphone should be attachable to the Exobreather (EB). The speakers should preferentially be integrated into the Exobreather and for that reason we wearable below the clothing. The on off switch and volume control could be controlled by a pressure sensor in the stoma.
The current available hands free systems work by giving an air pressure shock, closing the stoma. This system works quite intuitively in practise. In the client case the speech pressure range starts from 150mm water column and at 700 mm water column the clients speaks at full capacity.
The system does not need to be on at all times and should jump in when the user raises their voice. A suitable volume control is to be determined.

G.2. Statement of requirements

General requirements

- The System comprises of and suitable components are selected for:
  - A microphone
  - A signal processing part
  - An amplifier
  - A speaker
  - An energy source
- It is wearable on the body
- Due to the current corona health crisis a functioning prototype is not part of the SOR
- Its operating temperature stays below $36^{\circ}C$
- The input bandwidth is at least 50 Hz to 4 kHz
- The output sound intensity is at least 85 dB at 0.3 m
- The output bandwidth is at least 50 Hz to 4 kHz
- The microphone can not be a headset
- The volume control is hands free
- The system should provide 2 hours of speaking time,
- of which 30 minutes at high sound intensity
- The device should be rechargeable,
- with maximally 8 hours of charging time

Signal processing requirements

- Background noise is suppressed by 6 dB
- Acoustic feedback should be avoided
- The F_0 should be shifted up to resemble a less pathological voice
- The voice produced should be intelligible

Additional preferences

- The total costs of the parts should be around 100 euros
- It is integratable with the 'Exobreather'
- The size will have to be as limited as possible
- The latency of sound produced should stay below 15 ms
- The reproduced voice should be as natural as possible
- The volume should be regulated by air pressure

Closing statement

At the end of the project trajectory an extensive specification of the voice enhancement device / voice correction device is delivered. Another research group should be able to continue the research using the results of this preliminary investigation.
An NDA is part of the assignment.

G.3. Original statement of requirements

For completeness the original SOR is included below in Dutch.

Opdrachtformulering voor BAP studenten groep I

05-05-2020, Delft

Opdracht omschrijving:

Stem versterking voor mensen zonder stembanden:
De stemversterker maakt deel uit van een 2de generatie hulpmiddel voor LP Exobreather genaamd
Dit apparaat verzorgt het conditioneren van de lucht (reinigen, opwarmen en bevochtigen) dat anders in je neus mond keelholte gebeurt. Tevens komt er een voorziening zodat je handsfree kan spreken en er komt een slijmvanger zodat je maar een keer per dag je stoma etc hoeft te verzorgen. Dit apparaat wordt onder de kleding onder de stoma op het bovendeel van de borst gedragen. Het spraakversterker systeem maakt inclusief alle accessoires onderdeel uit van de Exobreather. Daarom moet alles zo compact mogelijk worden gebouwd.
We gaan voor een all-in one systeem dus bij voorkeur geen losse delen.
We gaan daarom voor een microfoon die dus aan de Exobreather (EB) bevestigd kan worden.
De speakers moeten bij voorkeur ook in de Exobreather worden ondergebracht die verdwijnen bij voorkeur dus ook onder de kleding. Het aan en uit schakeling en de volume control zou dmv een druksensor die in de stoma zit kunnen geschieden.
De handsfree systemen die je nu kan gebruiken werken namelijk als volgt: je geeft even met je adem een drukstoot dan sluit een klep de stoma af. En in de praktijk werkt dit behoorlijk intuitief.
In mijn geval kan ik beginnen met praten bij circa 150 mmWK en bij 700 mmWK presteer ik maximaal.
Het systeem hoeft niet altijd aan te staan het moet bijspringen zo gauw als je je stem wil verheffen er wordt nog een gepaste volumeregeling bepaald.

Programma van eisen

- Algemeen
    - Het systeem bestaat uit een microfoon, signaal bewerk systeem, versterker, speaker, energiebron, te dragen op het lichaam
    - Het leveren van een werkend prototype hoort vanwege de corona niet meer tot de opdracht.
    - De bedrijfstemperatuur blijft beperkt tot 36 °C
    - Geluid input spectrum minimaal 50Hz tot 4000Hz
    - Maximaal uitgangsgeluid minimaal 85dB at 0.3 meter
    - Frequentie bereik uitgang systeem 50Hz tot 4000Hz
- Microfoon
    - Selecteren van de microfoon
    - Geen headset
- Signaal bewerk systeem
    - Filter achtergrond geluid 6 dB
    - Filter rondzingen
    - Verschuif de toonhoogte naar die van een gezonde stem (statisch / dynamisch)
    - Stemgeluid moet verstaanbaar zijn
- Speakers
    - Selecteren van de speaker
- Versterker

    - Selecteren van de versterker
- Aan/uit en volume regel systeem
    - Volume regeling is hands free
- Energie bron
    - Opstellen van specificaties voor energiebron (inschatten van opslagcapaciteit)
        - Levensduur systeem = 2 uur op een dag spreken waarvan 30 min hard
        - Oplaad tijd = 8uur
    - Selectie van het meest geschikte type energiebron

Additionele wensen

- De totaalkosten richtlijn zal 100 euro bedragen
- Systeem wordt integreerbaar met de exobreather
- Het formaat blijft zo beperkt mogelijk
- Latency 15 ms
- Stemgeluid moet natuurlijk blijven
- Volume regelbaar op luchtdruk

AAN HET EIND VAN JULLIE ONDERZOEK LEVER JE EEN TO THE POINT UITGEBREIDE SPECIFICATIE VAN DE STEMVERSTERKER/STEMCORRECTIESYSTEEM WAAR DE “AFDELING” NA JULLIE WEER MEE VERDER KAN MET DE RESULTATEN VAN JULLIE VOORONDERZOEKINGEN

Een NDA is onderdeel van de opdracht.